

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 11, 2003, 11:55:44 ; Search time 39 seconds

(without alignments)  
3739.250 Million cell updates/sec

Title: US-10-009-782-1

Sequence: 1 gaattccactgacgcgcga.....ccctgacgacgagaagcct 1758

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=y1p  
-O/-CGN2.1/USPOT.SPOT/US10009782/runat\_07052003\_122517\_23133/app\_query.fasta\_1.1927  
-DB=SwissProt\_40 -QMT=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-ORIT=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdt -LIST=45  
-DOCALIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10009782.ecgn.1.1.19.8runat.07052003\_122517\_23133 -MCP=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGEDUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120  
-NARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELDP=6 -DELEXT=7

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2506	76.0	483	1	NDAD_ALCXX
2	1363.5	41.3	498	1	NDAD_ALCXX
3	1000.5	30.3	488	1	NDAD_ALCXX
4	321.5	9.7	660	1	YH1L_EBV
5	312	9.6	660	1	YH1L_EBV
6	296.5	9.0	1367	1	AMTH_YEAST
7	289.5	8.9	611	1	YH1L_EBV
8	284.5	8.8	5179	1	MHC2_HUMAN
9	284.5	8.6	1733	1	VND4_PRRNA
10	282.5	8.6	437	1	YGY3_HALSO
11	274	8.4	1733	1	VND4_PRRNA
12	270	8.2	1464	1	CA11_HUMAN
13	269.5	8.2	1027	1	CA11_HUMAN
14	268.5	8.1	1460	1	CA11_CANFA
15	267	8.1	825	1	5S5_RAT
16	264.5	8.0	1466	1	CA13_HUMAN
17	264	8.0	1464	1	CA13_MOUSE
18	263	8.0	1453	1	CA11_MOUSE

19	261	7.9	779	1	CA11_BOVIN	P02453 bos taurus
20	259.5	7.9	1453	1	CA11_CHICK	P02457 gallus gall
21	259	7.9	1255	1	MOC1_HUMAN	P15941 h. mactr 1 p
22	252.5	7.7	1459	1	CA12_MOUSE	P28481 mus musculu
23	244.5	7.4	1418	1	CA12_MOUSE	P02458 homo sapien
24	244	7.4	1461	1	IE18_HUMAN	P1675 pseudorabie
25	243.5	7.4	1372	1	CA21_MOUSE	O01149 mus musculu
26	242	7.3	555	1	GP1_CHLRE	O9F06 chlamydomon
27	240.5	7.3	1262	1	CA13_CHICK	P12105 gallus gall
28	239.5	7.3	1355	1	CA21_RANCA	O42350 rana catesb
29	239	7.2	1446	1	IE18_PRRNA	P33479 pseudorabie
30	239	7.2	1806	1	CA1B_HUMAN	P12107 homo sapien
31	238.5	7.4	825	1	SE5_RAT	O63003 rattus norv
32	238.5	7.2	1049	1	CA17_HUMAN	P04258 bos taurus
33	238.5	7.2	2944	1	CA17_HUMAN	O02388 homo sapien
34	237.5	7.2	671	1	EXTR_TORAC	P02454 rattus norv
35	237	7.2	620	1	CA2B_MOUSE	P13983 nicotiana t
36	235.5	7.1	1650	1	NME4_MOUSE	O64739 mus musculu
37	234.5	7.1	1323	1	CA21_BOVIN	O03391 mus musculu
38	234.5	7.1	1364	1	CA21_BOVIN	P02465 bos taurus
39	233.5	7.1	1372	1	CA1B_MOUSE	P02466 rattus norv
40	233	7.1	1804	1	CA21_MOUSE	O61245 mus musculu
41	232.5	7.0	1362	1	CA21_CHICK	P02467 gallus gall
42	231.5	7.0	2424	1	CCAA_RABIT	P27884 oryctolagus
43	231	7.0	448	1	MY15_HUMAN	O9176 delonococcus
44	231	7.0	3530	1	CA21_MOUSE	O9176 delonococcus
45	229	6.9	1366	1	CA21_CANFA	O46332 canis fam11

## ALIGNMENTS

RESULT 1  
ID NDAD\_ALCXX STANDARD; PRT; 483 AA.  
AC P23349: 008051:  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE D-aminocyclase (EC 3.5.1.81) (N-acyl-D-amino-acid deacylase).  
GN DAN.  
OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).  
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
OC Achromobacter;  
OX NCBI\_Taxid=515;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=96100942; PubMed=8541651;  
RA Wakayama M., Katsuno Y., Hayashi S., Miyamoto Y., Sakai K.,  
RT Moriyuchi M.;  
RT Cloning and sequencing of a gene encoding D-aminocyclase from  
RT Alcaligenes xylosoxydans subsp. xylosoxydans A-6 and expression of  
RT the gene in Escherichia coli.  
RL Biosci. Biotechnol. Biochem. 59:2115-2119(1995).  
CC -1- FUNCTION: HAS A WIDE SPECIFICITY; HYDROLYSES N-ACYL DERIVATIVE OF  
CC NEUTRAL D-AMINO ACIDS.  
CC -1- CATALYTIC ACTIVITY: N-acyl-D-amino acid + H(2)O -> an acid + D-  
CC amino acid.  
CC -1- COFACTOR: ZINC.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
EMBL: S80683; AAB35881.1; -  
EMBL: D45918; BAB08349.1; -  
KW Hydrolase; Zinc.

QY	937	GGAGGACCATATACATACCGTGGGCAAGGCGTCCCGCAACGTAGCGGGCGCGACCTGGAT	996
Db	301	GJAYgthrllellelthrttPcyslsyProPheProGluLeuSerGlyhAghApleAasp	320
QY	997	GAAGTCGGCGCGGACGGCGGCAATTCAAAGTACGACGTGTGTGCCGAGCTGCAGCGCGCC	105
Db	321	GIUValAlaAlaAGlAaGlyLylySerLystrYnAspValPProGluLeuGlnProAla	340
QY	1057	GGCGGCACTTACTTATATATGAGCAACCCAGCTGCAGCGCATCTGGCGTTGGCCCG	111
Db	341	GIYAlAlleTyYPhmeMetAspLupProAspValGlnArgIleLeuAlaPheGlyPro	360
QY	1117	ACCATGATCGGCTCCGAGCGGCTCGCGGACGACGAGCGCGCGCATCCGCGCTGTGGGGC	117
Db	361	ThrmelIleGlySerAspGlyLeuProHIsAspLunArgProHIsProAlaGleuTrpGly	380
QY	1177	ACCTTCGGCGGGGTCTGGGCACTATGCGCGGACACTGGGCGCTTTCCCGTGGAGACG	123
Db	381	ThrPheProAValLeuGlyhIstYrAlaArgAspLeuGlyLeuPheProLeuGlnTrh	400
QY	1237	GGCGATGTGAGATGACCGGCGCTGACCGCGCGCGCTTGGCGCTGGCGGGCGGGGCAg	129
Db	401	AlaValTrpLysMetThrGlyLeuThrAlaAlaArgPheGlyLeuAlaGlyhGlyGln	420
QY	1297	CTGCAGGCGGGGTACTTGGCGGACCTGGGTGTTCGACCCGGCGACGATGGCGGATACC	135
Db	421	LeuGlnAlaGlyTyYrPheAlaAspLeuValAlaPheAspProAlaThrValAlaAspTrh	440
QY	1357	GGCACCCTTGAAACACCTTACCGAGCGGGCGCGCGGCGATTCGATCCGTACGTCAACGGC	141
Db	441	AlaThrPheGlnHIsProThrGlnAlaAlaAlaGlyIleHIsSerValTyYrAlaGlnGly	460
QY	1417	GGCGCGGTCTGGCAAGACAGCGCTTACCGCGGCGAGCATCCGCGCGTGGTGCAGCGC	147
Db	461	AlaProValTrpGlnGlnAlaPheThrGlyGlnHIsAlaGlyhArgValLeuAlaArg	480
QY	1477	ACGGCGCGCC	1485
Db	481	ThrlAlaAla	483
RESULT 2			
	NDND_ALCXX	STANDARD;	PRT; 498 AA.
ID	NDND_ALCXX		
AC	P94212;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUL-1998 (Rel. 36, Last sequence update)		
DT	15-JUL-1998 (Rel. 36, Last annotation update)		
DE	N-acyl-D-aspartate deacylase (EC 3.5.1.83) (N-acyl-D-aspartate		
DE	amidohydrolase).		
OS	Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxidans).		
OC	Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;		
OC	Achromobacter.		
OX	NCBI_TaxID=515;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A-6;		
RX	MEDLINE=3372486; PubMed=77633985;		
RA	Moriyuchi M., Sakai K., Katsuno Y., Maki T., Wakayama M.;		
RT	"Cloning, expression, and nucleotide sequence of the N-acyl-D-		
RT	aspartate amidohydrolase gene from Alcaligenes xylosoxydans subsp.		
RT	xylosoxydans A-6".		
RL	J. Ferment. Bioeng. 80:311-317(1995).		
RL	[2]		
RP	CHARACTERIZATION.		
RC	STRAIN-A-6;		
RX	MEDLINE=3372486; PubMed=77633985;		
RA	Moriyuchi M., Sakai K., Katsuno Y., Maki T., Wakayama M.;		
RT	"Purification and characterization of novel N-acyl-D-aspartate		
RT	amidohydrolase from Alcaligenes xylosoxydans subsp. xylosoxydans		
RT	A-6".		
RL	Biosci. Biotechnol. Biochem. 57:1145-1148(1993).		
CC	-1- CATALYTIC ACTIVITY: N-acyl-D-aspartate + H(2)O = carboxylate + D-		
CC	aspartate.		





CC -1- COFACTOR: ZINC.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC EMBL: D50061; BAA08778.1;  
 DR InterPro: IPR002195; Dihydrocoarctase.  
 DR Pfam: PF00744; Dihydrocoarctase; 1.  
 KW Hydrolase; Zinc.  
 SQ SEQUENCE 488 AA; 51492 MW; F5A1B8315BEA167F CRC64;  
 Alignment Scores:  
 Pred. No.: 1,93e-35 Length: 488  
 Score: 1000.50 Matches: 223  
 Percent Similarity: 59.38% Conservative: 62  
 Best Local Similarity: 46.46% Mismatches: 190  
 Query Match: 30.33% Indels: 5  
 DB: 1 Gaps: 4  
 US-10-009-782-1 (1-1758) x NDED\_ALCX (1-488)  
 QY 52 CAGCCCTTCGACCTGCTGCGGGGCGACCCCTCATGAGCGAGCAACACCCGGG 111  
 DB 3 GILVSLAASPLEUVALLEGLUNGLYTPVAILLEASPGLYLEUGLYLPROARG 22  
 QY 112 CGCGCGCGCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 171  
 DB 23 AGAGALAAASPVAILGLYLGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLY 42  
 QY 172 GCGCGCGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 231  
 DB 43 ALAPROALAAAPARARGLEUASPAALAGLYARGILEVALAIPROGLYPHEILEASP 62  
 QY 232 TCGCACACCCAGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 291  
 DB 63 THIRHSLGLYHSLASPSAPLEUMETHEVALGLULYSPROGLYLEUGLITRIPLYSHTSER 82  
 QY 292 CAGGCGCGTACACGCGTGTACAGCGCGCAATTCGCGCGCGCGCGCGCGCG 348  
 DB 83 GINGLYLITHTSERVALVALVALGLYASNCYGLYILSERGLYALAPROALPROLEU 102  
 QY 349 CAGCGCAACCG 408  
 DB 103 PROGLYASHTHRLAALALALEUALALEU-----GLYASPSERPROLEUPHEALA 120  
 QY 409 CGCTTCGCGCACTACCTGACGCGCTTCGCGCGCGCGCGCGCGCGCGCGCG 468  
 DB 121 ASPMETALMETTYRPHGLYALALEUUNLALAGLNAIPROMETILLEASVVALAALA 140  
 QY 469 ATGGGCGCGCATTCACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528  
 DB 141 LEUVALGLYHSLASPSAPLEUVALLEGLUNGLYTPVAILLEASPGLYLEUGLY 160  
 QY 529 GAGCGAGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 588  
 DB 161 ALALYSGULINARGALAMEGLUARGLEUVALLEASPAALAGLUNGLYALVAL 180  
 QY 589 GGCATTTTCGACCGCGCGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCG 648  
 DB 181 GLYPHESERTHRLYLEUALATYGLINPROGLYLYALALAGLUNGLINLEUASP 200  
 QY 649 GAGGTGCG 708  
 DB 201 GLYLEUAAARGVALAALAAALAAARGLYALALEUHTHTSERHSLILEARGASGLU 220  
 QY 709 GCGGAGACATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 768

DB 221 GLYASPAALVALGLUVALAALVALASPGLEUVALLEUALVALGLYARGARGTHRGLYCYS 240  
 QY 769 CCGGTGTATATTCGACCAACAGATGAGCGCGCGCGCGCGCGCGCGCGCG 828  
 DB 241 ALATHVALLEUSERTHSLISLSCYSMETMETPROALASHTPGLYASERIALA 260  
 QY 829 AGCGTCCCGCGTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885  
 DB 261 THIRLEALASHTHLEASPSARGVALAARGALAAALAGLYVALASPVALLALEUASPLETYR 280  
 QY 886 CCGTACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945  
 DB 281 PROTYRPROGLYSESTERTHRLLEULEUETPROGLUARGALASPGINLEASPSALLE 300  
 QY 946 ATCATCACCCTGTGCAAGCCCTTCCCGCACTGACGCGCGCGCGCGCGCG 1005  
 DB 301 ARGILETHRTPSERTHPRONHISPROGLUCYGLYGLINSEUVALLEGLUVAL 320  
 QY 1006 GCGGAGCGCGCAATTCAGAGACGAGTGGCGCGCGCGCGCGCGCGCG 1065  
 DB 321 ALAARGTPGLYCYSASPAALVALTHRLAALAAARGARGLEUCYSPROGLYALALE 340  
 QY 1066 TACTTCATGATGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1125  
 DB 341 TYRPHLEALAMEASPSGLUASNGULVALARGARGILEPHEGLINHLSGLUCYSCYSMETVAL 360  
 QY 1126 GCGTCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1185  
 DB 361 GLYSERASPSGLYLEUPROASNSPALHSLSPROHISPROARGLEUTRIPLYSERPHETHR 380  
 QY 1186 CGGGTGTGGGCGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1245  
 DB 381 ARGVALLEUGLYARGTYVALARGGLNALAGLEULEUTHRLLEUGLVALAALVALA 400  
 QY 1246 AAGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1305  
 DB 401 LYSMETHTHRLALEUPROALAAARGVALPHEGLYLEUVALAASPARAGLYARGLEUVALA 420  
 QY 1306 GGGTACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1365  
 DB 421 GLYALERTPALAASPVALLVALAALPHEASPSALASPTHVALCYASPARGLAATHTYR 440  
 QY 1366 GAACACCTTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1425  
 DB 441 ASPALAPROHTHRLLEUALASERIALGLYILEUHTHSLVALLEUVALASNGLYCYAALVAL 460  
 QY 1426 TGCGAAGAGCAGCGCGTTCACCGCGCGCGCGCGCGCGCGCGCGCG 1485  
 DB 461 PHE---PROGLNALAPROPSERTHSLARGPROGLYARGILEUARGALARGASPAALASER 479  
 RESULT 4  
 YHLL\_EBV  
 ID YHLL\_EBV STANDARD; PRT; 660 AA.  
 AC P03181;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Hypothetical BHFL1 protein.  
 OS Epstein-Barr virus (strain 95-8) (Human herpesvirus 4).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 OX NCBI\_TaxID=10377;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84270667; PubMed=6087149;  
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,  
 RA Tufnell P.S., Barrett B.G.;  
 RT "DNA sequence and expression of the 95-8 Epstein-Barr virus genome.";  
 RC Nature 310:207-211 (1984).  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch).)

CC EMBL: V01555; NOT\_ANNOTATED\_CDS.  
 DR PIR: A03742; Q0BE3.  
 KM Hypothetical protein; Early protein; Repeat.  
 FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.  
 FT REPEAT 149 273 1.  
 FT REPEAT 274 398 2.  
 FT REPEAT 399 523 3.  
 FT REPEAT 524 648 4.  
 SQ SEQUENCE 660 AA: 66244 MW: 86DA1D67A37152A2 CRC64;

## Alignment Scores:

Prod. No.: 4.436-07 Length: 660  
 Score: 321.50 Matches: 187  
 Percent Similarity: 31.83% Conservatve: 18  
 Best Local Similarity: 29.04% Mismatches: 236  
 Query Match: 9.75% Indels: 183  
 DB: Gaps: 29

US-10-009-782-1 (1-1758) x YH1\_EBV (1-660)

QY 39 CCAATCCGATCCGACCCCTTCGACTGCT----- 68  
 DB 113 ProGlnArgSerProArgThrArgGlnAlaGlyTyrAlaLeuGlyGlySerAlaGly 132  
 QY 69 -----GCTGGGGGGGGGACCCCTCATCGACGCGACACACCCGGGGGGGGCCGCA 122  
 DB 133 LenglySerAlaGlyProArgProAlaPhaeGlnAlaGlnTrpSerAlaGlnAsn 152  
 QY 123 CTTGGGCGTGGCGGGGACCGCATCGCGCATCGCATCTGCGAGCGCGCGCGCA 182  
 DB 153 ProGlyCysPro-----ArgThrTrpArgArgArgSerGlyAlaGlnAlaGly 168  
 QY 183 CACCGGGGTGACGCTGCGGGCTGTGGTGGCGCGCGCTCATGACTGCACACCA 242  
 DB 169 HisPro----- 170  
 QY 243 CGACGACAACTACCTGCTCAGCGGTGGCGACATGACGCCCAAGATCTCGAGGGCGTAC 302  
 DB 171 -----ProProGlyAlaGlyGlnArgProSerGlyProThrGlyArg--- 185  
 QY 303 CACGGTGTACACGGCAATTCGGCATCAAGCGCGCGCGCGCGCGCAACCGCGC 362  
 DB 186 -----ProAlaAlaProGlyAlaProGlyThrPro 195  
 QY 363 CGCCCGCTGGACCTGCTGAGCAGAGCGGCTTACCGTTTCGAGCGCTTCGCGCACTA 422  
 DB 196 AlaAlaProGlyProGlyGly---GlyAlaAlaValPro---SerGlyAlaThrProHis 214  
 QY 423 CCTGGACGGCTTCGGGCA-----CGCC 446  
 DB 214 roGlnArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlnArg 234  
 QY 447 GCGGCGCTTCACGCGCGCTGTATGGTGGGCGCATTCACGCTGCGCGCGCGGTGATGCG 506  
 DB 234 InGluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysPro----- 249  
 QY 507 GGAATTGACAGCGCGCGCACGACGAGAGAAATCGCGGCGCATCGGGACCTGGCCGAGA 566  
 DB 250 -----AlaGlyProProProThrArgSerGly----- 258  
 QY 567 AGCGATGGCAGCGGGCGCATCGGCATTTTGACCGCGGCTTCTACCCCGCGCGCGCG 626  
 DB 259 -----AlaAlaAlaGlnArgThrHisArg-----ArgProProGlyC 271  
 QY 627 CGCGACACCGAGAGATCATCGAGGTGTCGCGCGCGCTGAGCGCGCATGCGCGCATCTA 686  
 DB 627 ----- 686

DB 271 ySPPro-----ArgSerAlaArg-AsnProGlyCysProArgThrTrpArg----- 285  
 QY 687 CGCCACCCATCGCGACGAGAGCGGACACATCTGGCGCG-----GCTGAGAGA 737  
 DB 286 -----ArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGly 300  
 QY 738 AACCTTCGCATCGCGCGCGGACCTGAGCTCCGGTGTGATCTGCACACAAAGTCAT 797  
 DB 301 GlnArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaPro----- 317  
 QY 798 GGGCGACCCCATTTTCGGCGCGCTCGCGGAGACGCTGGCGCGATCGAGCGCGCATGCGC 857  
 DB 318 -----GlyThrProAla 321  
 QY 858 GCGCGAGACGCTGCTGCGAGCGGATTCCTACCTGCGCGCGCGCTCCACCATCTACACA 917  
 DB 322 AlaProGlyProGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisProGlnArg 341  
 QY 918 GGA-----CGCGTGTCTGCGCGGACGCGACGACGACATCATCTGCTGCA 962  
 DB 342 GlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlyAlaGlnArg 361  
 QY 963 GCGCTTCGCGGACGCGGCGCGGACCTGGATGAACTCGCGCGCGGCGGCGGCAATC 1022  
 DB 362 ArgLeuProGlnAspLeuAlaAla-----GlnArgCysProAlaGlyProPro 378  
 QY 1023 CAAGTACAGCTGTGTCGCGGACCTGACGCGCGCGCGCGCATCTTACTTACATGAGAGCA 1082  
 DB 379 ProThrArgSerGly-----AlaAlaAlaGlnArgThrHisArgArgProProGlyCys 396  
 QY 1083 ACCCGACGTGACGCGGATCTGCGCGCTGGCGCGGACATGATCGGCTCGGACGCGCTGCGC 1142  
 DB 397 ProArgSerAlaArgAsnProGlyCys---ProArgThrTrpArgArgSerGlyAla 415  
 QY 1143 GCACGACGAGCGCC-----GCATCCGCGCTGTGGGCGACCTT 1181  
 DB 416 GlnArgGlyHisProProProGlyAlaGlnArgProSerGlyProThrGlyAlaArg 435  
 QY 1182 CCGCGGGTGTGGGCA-----CTATCGCGCGCATCTGGCGCT----- 1220  
 DB 436 ProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAla 455  
 QY 1221 GTTCCCGCTGGA-----GACGGGCGTATGAGATGACGCGCGCTGAC 1262  
 DB 456 ValProSerGlyAlaThrProHisProGlnArgGlySerGlyProAlaAspProProAla 475  
 QY 1263 CGCGCGCGCTTCGCGGCGCGCGCGCG-----GCACCTGCA 1301  
 DB 476 AlaAlaArgLeuProProGlyAlaGlnArgProArgLeuProGlnAspLeuAlaAla 495  
 QY 1302 GCGCGGTACTTCGCGGACCTGTGTGTCACCGCGCGCGCGCGCGCATACCGCGAC 1361  
 DB 496 GlnArgCysProAlaGlyProProProProThrArgSerGlyAlaAlaGlnArgThrHis 515  
 QY 1362 CTTCGACACCCCTACCGGCGCG-----CGCGGACATCTTCGCTGATCGTCAAGC---G 1415  
 DB 516 ArgArgProProGlyCysProArgSerAlaArgAsnPro---GlyCysProArgThrTrp 535  
 QY 1416 CCGCGGCTTCGCGACGAGCGGCTTCACGCGGACGACATCGCGCGCGCGCTGCGACG 1475  
 DB 535 rGlnArgSerGlyAlaGlnArgGlyHisPro----- 545  
 QY 1476 CACGCGCGCTGAGCGCGCGCGCGCGCTTACATCGCGCGGTGAAAGGCGCGCGCTGCGC 1535  
 DB 546 -----GlyAlaProProAla 563  
 QY 1536 CCGCGCGCGCGGACGCGGCAACCGCTACATGCGCC-----CTCCCTCGCG 1583  
 DB 563 IaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProS 583  
 QY 1584 TCGCAATTCAGCGCGCGCGCATATCTGCGGCAAGAAATGATGGCGCGCGCTGCGCGC 1643  
 DB 583 erGlyAlaThrProHisPro-GlnArgGlySerGlyProAlaAspProProAlaAlaAla 602

QY 1644 CGAGCGCAAGCCCGGAAATGACCCCTGCAGACCT-----GTGCGAGCCGACGCG 1694  
 DB 603 ArgLeuProProGluArgGlnGlnIupProArgLeuProGlnAspLeuAlaAlaGlnArg 622  
 QY 1695 ---CATGCGGCTCTGAGCCAGCTGCAGCCGAGTGGCCAGATGCGCTGAGTACGA 1751  
 DB 63 CysProAlaGlyProProProthThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArg 642  
 RESULT 5  
 YHL1\_EBV  
 ID YHL1\_EBV STANDARD: PRT: 660 AA.  
 AC P03181;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Hypothetical BHLF1 protein.  
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 CX NCBI\_TaxID=10377;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84270667; PubMed=6087149;  
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,  
 RA Tufnell P.S., Barrett B.G.;  
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome";  
 RL Nature 310:207-211(1984).  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: V01555; .; NOT\_ANNOTATED\_CDS.  
 DR PIR: A03742; Q0BE3.  
 KM Hypothetical protein; Early protein; Repeat.  
 FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.  
 FT REPEAT 149 273 1.  
 FT REPEAT 274 398 2.  
 FT REPEAT 399 523 3.  
 FT REPEAT 524 648 4.  
 SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;  
 Alignment Scores:  
 Pred. No.: 1,1e-06 Length: 660  
 Score: 312.00 Matches: 181  
 Percent Similarity: 31.59% Conservative: 30  
 Best Local Similarity: 27.10% Mismatches: 232  
 Query Match: 9.62% Indels: 225  
 DB: 1 Gaps: 31  
 US-10-009-782-1 (1-1758) x YHL1\_EBV (1-660)  
 QY 1753 TCTGTAGCTCAGGCGCATCTGCCACAGT-----CGGCGT 1718  
 DB 87 SerArgGlnSerArgThrGlnProAlaGlnAlaAlaSerHisSerAsnPro 106  
 QY 1717 TGCACAGGTCGACAGCCGATCCGCTGCGTCGACAGCTTTCGAGGTCATTTCC 1658  
 DB 107 ThrGlyGlyCysSerAspProGlnArgSerProArgThrArgGlnAlaGlyTyrAlaLeu 126  
 QY 1657 GGGCCCTTGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1598  
 DB 127 GlyGlnGlySerAlaGlyLeuGlySerArgGlyProArgProHisProAlaPheGlnVal 146  
 QY 1597 -----GGGCGGTATTC-----1586  
 DB 147 GlnTrpSerAlaArgAsnProGlyCysProArgThrTrpArgArgArgSerGlyAlaGln 166

QY 1585 -----GAGCGGAGGAGGAGGCGCATGACGGT----- 1559  
 DB 167 ArgGlnHisProProProGlyAlaGlyAlaGlnArgProSerGlyProThrGlyArgPro 186  
 QY 1558 -----TTGCGTCCAGGAGGTTGGAGGAGGCG----- 1535  
 DB 187 AlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaVal 206  
 QY 1534 -----GCGACGCGGCGCGGTCACGCGGATGTGAGGCGTGGCGCGCGCTGAGCGG 1481  
 DB 207 ProSerGlyAlaThrProHisProGlyArg-----GlySerGlyProAlaAsp 222  
 QY 1480 CCGTGGCGTGCAGACGCG-----GCGCGG 1457  
 DB 223 ProProAlaAlaAlaAlaAlaGlnLeuProProGlyAlaGlnIupProArgLeuProGlnAspLeu 242  
 QY 1456 CATGCTGGCGCGGTACACGCTCTCTTCCAGACGCGG----- 1418  
 DB 243 AlaAlaAlaGlnArgCysProAlaGlyProProProthThrArgSerGlyAlaAlaGln 262  
 QY 1417 CGCGCTTGACGTACAGGATGATGCCGCGCGCGCTCGTAGGT-----GTTCGA 1364  
 DB 263 ArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArg 282  
 QY 1363 AGTGGCGGTATTCG-----CCACGCTGGCGCGGTGCAGACA 1328  
 DB 283 ThrTrpArgArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyAlaArg 302  
 QY 1327 CCACGAGTGCAGGAGTACCGGCTGACCT-----GCCGCGCGCGG 1283  
 DB 303 ProSerGlyProThrGlnGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAla 322  
 QY 1282 CCAGCGCGGAGCGCGCGCGCGCTGACGCGGCTTCATCCGCGCTCCAGCGGA 1223  
 DB 323 ProGlyProGlyGlyGlyAlaAlaAlaValProSerGlyAlaThrProHisProGlyAlaGly 342  
 QY 1222 ACAGCGCCAGGTCGCGGCGCATAGTGCCTCA-----1193  
 DB 343 SerGlyProAlaAspProProAlaAlaAlaAlaArgLeuProProGlyAlaGlnIupProArg 362  
 QY 1192 -----GCACCGCGGAGAGTGCCTCCAGCGCGGATGCGGCGCTGCT 1148  
 DB 363 LeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProthArgSer 382  
 QY 1147 CGTGGCGACGCGCTGCGGACCGCATCATGCTGGCGCGGAGCGGACGCGGCTGCGACGT 1088  
 DB 383 GlyAlaAlaAlaGlnArgThrHisArg-----ArgProProGlyCysProArgSer 399  
 QY 1087 CGGGTTCGTCATCATGAGTAGA---TGCGCGCGCGCGGCTGCGGACGCGGCTGCGACGT 1031  
 DB 400 AlaArgAsnProGlyCysProAlaGlyThrTrpArgArgArgSerGlyAlaGlnArgGlyHis 419  
 QY 1030 CGTACTTGATTTGGCGCGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 971  
 DB 420 ProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaPro 439  
 QY 970 GGAAGGCTTGACACAGGTATATGCTGCTGCGGCGGCGGCGGCGGCGGCTGCTGCT--- 914  
 DB 440 GlyAlaProGlyThr-----ProAlaAlaProGlyProGlyGly 452  
 QY 913 -----TGACATGCTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 869  
 DB 453 GlyAlaAlaValProSerGlyAlaThrProHisProGlyAlaGlySerGlyProAlaAsp 472  
 QY 868 CGTCTGCGGCG-----CCATGCGGCGCTGCATACGCG 836  
 DB 473 ProProAlaAlaAlaAlaArgLeuProProGlyAlaGlnIupProArgLeuProGlnAspLeu 492  
 QY 835 GAGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 776  
 DB 493 AlaAlaAlaGlnArg-----CysProAlaGlyPro 502

```

OY 775 CCAACGGCAGCT-----CCAGCTCGGGCGGCGATGCGGAAGTTTCTCCAGCGCGGCCA 722
DB 503 ProthrinArgSerGlyAlaAlaAlaGlnArgHisArgAArgProGlyCysPro 522
OY 721 CGATGCTCGCCTTCCTCCGCGATGTGGTGCGTAGATGCCGATGCCGCTCAGCG 662
DB 523 ArgSerAlaArg-----Asn 527
OY 661 GCGGCGACACCTCATATCTCTTCGTTGGTGGCGGGCGGGCGGGTAGAGGCCG 602
DB 528 ProGlyCysPro-----ArgThrTrpArg---ArgArgSerGlyAlaGlnArg 542
OY 601 CCGGCGAAATGCGGATGCGCGCGCGGCGCATGCTTCCTCGCGGAGCGCGCGATGCGG 542
DB 543 -GlyHisProProGlyAla-----GlyGlnArgProSerGlyPr 556
OY 541 CGATTCCTCGTGCGTGCGCGCGCGCTCAGATGCCATGCGCGCGCGCGCGCTTG 482
DB 556 o-----ThrGlyGly-ArgProAlaAlaProGlyAlaProGlyThrProAlaAla 572
OY 481 AATGGCCCA-----CCATACAGCGCGCGCTTGA 455
DB 573 ProGlyProGlyGlyGlyAlaAlaAlaProSerGlyAlaThrProHisProGlyArgGly 592
OY 454 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 395
DB 593 SerGlyProAlaAspProAla-----AlaAlaAlaArgLeuProPro 606
OY 394 AGCGCGCTTCGTCGCGAGGTCGCGGGGGG-----CGGCGGGGTTGCGCGCGCGCGCG 338
DB 607 GlnArgGlnGlnProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGly 626
OY 337 CCAAGCGTGAATGCCCAATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 278
DB 627 Pro-----ProPro-----629
OY 277 TCATGTCGCGAGCGCTGAGCAGTAGTGTGCGTGCGTGCGCGATGAAGCGCG 218
DB 630 -----ThrArgSerGlyAlaAlaAlaGlnArgThrHis 640
OY 217 GCGCGACACCGACGCG-----CCGACACGTCGACCGCGCGTGCGCGCGCGCGCG 167
DB 641 ArgArgProProGlyCysProArgSerAlaArgAsnProGlyCys-----655
OY 166 ACAGATCGCGGATGCGCG 149
DB 656 ---ProArgThrTrpArg 660

RESULT 6
AMYL_YEAST STANDARD: PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucanase SI/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
glucosidase) (1,4-alpha-D-glucan gluconhydrolase).
GN STAI OR STAI OR MALD OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288c / AB972;
RA Barrer C.M., Badcock R., Bankier A.T., Bowman S., Brown D.,
RA Churchill C.M., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentes S., Hamlyn N., Horsnell T.S., Hunt S., Jagsels K., Jones M.,
RA Rajadream M.A., Riles L., Moulé S., Moulé T., Odell C., Pearson D.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STAI";
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Imanez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STAI and SGA genes
RT from Saccharomyces cerevisiae";
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S. POMBE SPBC215.13.
CC -1- SIMILARITY: SOME, TO S. POMBE SPCC285.13C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; Z38061; CA86176.1; -
DR EMBL; M16164; AAA35014.1; -
DR EMBL; M16165; AAA35015.1; -
DR EMBL; X13857; CA32069.1; -
DR PIR; B26877; B26877.
DR PIR; A26877; A26877.
DR PIR; S48478; S48478.
DR SGD; S0001458; MDCL.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 874
FT CARBOHYD 874 874
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CMC64;

Alignment Scores:
Pred. No.: 4.6e-06 Length: 1367
Score: 296.50 Matches: 139
Percent Similarity: 38.72% Conservative: 72
Best Local Similarity: 25.50% Mismatches: 266
Query Match: 8.99% Indels: 68
DB: Gaps: 13

US-10-009-782-1 (1-1758) x AMYL_YEAST (1-1367)
OY 8 ACTGATCGCGGAGAGAGATTCATGTCCTCATATCCATTCACCTTCGACCTGC 67
DB 345 ThrSerSerThrThrGlnSerSerSerAlaProValProThrProSerSerThr 364
OY 68 TGCCTCGGGGGGACCTATGACGCGACGACACCGCGGGGGCGGGCGGACCTGG 127
DB 365 GlnSerSerSerAlaProValThrSerSerThrThrGlnSerSerSerAlaPro 382
OY 128 GCGTGGCGGCGGCGCGCATGCGCGCGCGATGTCGTGAGCGCGCGCGCACACC 187
DB 383 ValThrSerSerThrThrGlnSerSerSerAla-----ProValProThrPro 398
OY 188 GCGTGCACGTCGCGCGCTGTGTGCGCGCGCGCTTCATCGACTGCACAA----- 238
DB 399 SerSerSerThrThrGlnSerSerSerSerAlaProValThrSerSerThrThrGln 418
OY 239 -----CCGACGAGAACTACTCTCTCAGCGCGTGCAGACGACCCCAAGATCGCG 292

```

```

Db 419 SerAlaProValThrSerSerThrThrgluserSerSerAla-----ProValThrSer 436
QY 293 AGGGCTGACACGAGGTGATACGGGCAATTGGCGATCAGGCGGCGCGCGCGCGCGCG 352
Db 437 SerThrThrgluserSerSerAlaProValThrSerSerThrThrgluserSerSerAla 456
QY 353 CCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 412
Db 457 ProValProThrPro-----SerSerSerThrThrgluserSerSer 470
QY 413 TCGCGG-----ACTACGTGGAGCGCGTGGCGGCGCGCGCGCGCGCGCGCGCGCG 460
Db 471 AlaProValThrSerSerThrThrgluserSerSerAlaProValProThrProSerSer 490
QY 461 CCGCGCTGATGTGGCGCATTCACGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 520
Db 491 SerThrThrgluserSerSerAlaProValThrSerSerThrThrgluserSerSerAla 510
QY 521 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 580
Db 511 ProValProThr----- 514
QY 581 GCGCGCATTCGCGATTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 640
Db 515 -----ProSerSerSerThrThrgluserSerSerAlaProValProThrProSerSer 533
QY 641 ACATCATCGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 700
Db 534 ThrglThrgl-----SerSerSerAlaProValThrSer 544
QY 701 GCGAGGAGGCGGAGACATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 760
Db 545 SerThrThrgluserSerSerAlaProValProThrProSerSerSerThrThrgluser 564
QY 761 TGGAGTCCGCGGTGTGATTCGCGACGACGAGCGCATGGCGCGCGCGCGCGCGCGCG 820
Db 565 SerSerThrProValThrSerSerThrThrgluserSerSerAlaProValProThrPro 584
QY 821 CCGCGGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 880
Db 585 SerSerSerThrThrgluserSerSerAlaProValProThrProSerSerSerThrThrgl 604
QY 881 CGTATCCCTAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 940
Db 605 GluserSerSerAlaProValProThrProSerSerSerThrThrgl 940
QY 941 GCACCATCATCATCT-----GGTGAAGCGCGTTCGCGCGCGCGCGCGCGCG 976
Db 621 SerSerSerAlaProValThrSerSerThrThrgluserSerSerAlaProValProThr 640
QY 977 TGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1036
Db 641 ProSerSerSerThrThrgluserSerSerAlaProValProThrProSerSerSerThr 660
QY 1037 TGGCGGAGTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1096
Db 661 ThrgluserSerSerAlaProValProThrProSerSerSerThrThrgluserSerSer 680
QY 1097 GCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1150
Db 681 AlaProValThrSerSerThrThrgluserSerSerAlaProValThrSerSerThrThrgl 700
QY 1151 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1210
Db 701 Gluser-SerSerAlaProValProThrProSerSerSerThrThrgluserSerSerAl 720
QY 1211 ACCTGGCGCGTGTCCCGCTGAGAGCGCGGTATGGAAGTACCGCGCGCGCGCGCGCG 1270
Db 720 AlaProValProThrProSerSerSerThrThrgluserSerSerAlaPro--ValProThr 739
QY 1271 GCTTGGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1330
Db 740 ProSerSerSerThrThrgluserSerSerAlaProValThrSerSerThrThrgluser 759

```

---

```

QY 1331 TCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1390
Db 760 SerSerAlaPro--ValProThrProSerSerSerThrThrgluserSerSerAlaPro 778
QY 1391 GCATTCATTCGCTGATCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1450
Db 779 ValProThrProSerSerSerThrThrgluserSerSerAlaProValProThrProSer 798
QY 1451 AGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1510
Db 799 SerSerThrThrgluserSerSerAlaProValProThrProSerSerSerSer--Am 817
QY 1511 CCGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1570
Db 818 IleThrSerSerAlaProSerSerThrThrgluserSerSerSerSerSerSerSer 837
QY 1571 CCGCTGCGT 1579
Db 838 ProValPro 840

```

RESULT 7  
 YTL3\_MYCTU STANDARD: PRT; 611 AA.  
 AC 010830;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein Rv2913c.  
 GN Rv2913c OR MT2981 OR MTCY274.45C OR MTCY338.01C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID:1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37Rv.  
 RX MEDLINE-98295987; PubMed-9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garner T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagers K., Krogan A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sluson J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Feldman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 DR EMBL; 274024; CA98380.1;  
 DR EMBL; AE007120; AA047307.1;  
 DR TIGR; MT2981;



DR Tuberculin: RV2913c;  
 KW Hypothetical protein: Hydrolase; Complete proteome.  
 SQ SEQUENCE 611 AA; 67205 MW; C3AA4A13B08BC34A CRC64;

## Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Best Local Similarity:	Mismatches:	Indels:	Query Match:
6,28e-06	611	294.00	153	79	24.064	216	190	1
					8.918		25	

US-10-009-782-1 (1-1758) x YP13\_MYCTU (1-611)

```

OY 58 TTCGACCTGCTGCGGGGCGGACCCCTCATGCGGCGACACACCCCGGGGCGGCC 117
DB 16 TTTASPVAILLEILLeuArgSPgLYLeuTPPhaSPgLYLeuTgYsnaLaProLeuThr 35
OY 118 GCCGACCTGGGCGTGGCGCGACCCGATCGCC-----GCCATGGCGATGTGTGGAC 171
DB 36 ArgThrLeuGlyLeuArgSPgLYLeuValAlaThrValAlaAlaGlyAlaLeuSPgLY 55
OY 172 GCCCGCGCGCACACCCGGGTGCGAGTGTGGCGGCGCTGTGCGCGCGGCTTCATCGAC 231
DB 56 ThrGlyCysProGluValAlaAspAlaAlaGlyLeuTyrValValProGlyPheLeuAsp 75
OY 232 TCGGACACCCGACGACACACACTGCTGACGCGTCCGCGACATGACGCCCAAGATCTCG 291
DB 76 ValHisThrHisTyrAspAlaGluValLeuLeuAspProGlyLeuArgGluSerValArg 95
OY 292 CAGGCGCTCACACGCTGTGTCACGGGCAATTGGCGCATCGCTGGGCGCGCGCGCGCAC 351
DB 96 HisGlyValThrThrValLeuLeuGlyAsnCysSerLeuSer-----ThrValTyr 112
OY 352 GCCAACCCGCGCGCGCGCGCTGCACTGCTGAGACGAAAGC----- 390
DB 113 AlaAsnSerGluAspAlaAlaAspLeuPheSerArgValGluAlaValProArgGluPhe 132
OY 391 -----GGCTCTTAACGT-----TTCGACCGCTTCGCCGACTACCTGAGCGG 432
DB 133 ValLeuGlyAlaLeuArgSPgLYLeuArgSPgLYLeuTyrPheThrProAlaGluTyrLeuGluAla 152
OY 433 TTGCGGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492
DB 153 IleAspAlaLeuProLeuGlyProAsnValSerSerLeuLeuGlyHisSerAspLeuArg 172
OY 493 GCCCGGCTCATCCCGGACTTGCAGCGCGCGCGC-----ACCGAGAG 534
DB 173 ThrAlaValLeu-----GlyLeuAspArgAlaThrAspAspThrValArgProThrGluAla 191
OY 535 GAAATCGCGCGCATGCGGAGACTGGCGCGGAGAAAGCATGCGCGCGCGCGCGCGCGCAT 594
DB 192 GluLeuAlaLeuMetAlaLysLeuLeuAspGluAlaLeuGluAlaGlyMetLeuGlyMet 211
OY 595 TCGACCGGCGCTTCTACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAT 647
DB 212 Ser-----GlyMetAspAlaAlaAlaIleAspLysLeuAspGlyAsp 224
OY 648 -----CGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 695
DB 225 ArgPheArgSerArgAlaLeuProSerThrPheAlaThrThrArgGlyLeuArgGlyLeu 244
OY 696 CATGCGCGCGACGAGCGAGCATGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 755
DB 245 IleSerValLeuArg-----HisArgGlyArgIle-----LeuInsAlaPro 259
OY 756 CGAGCTGACGTCGCGGTGTGATCTCGGACACCAAGTCATGGCGCGCGCGCGCGCGCGCG 815
DB 260 -AspValAspAsnProValSerAlaLeuLeuPhePheLeuAlaSerSerArgIlePheAs 279
OY 816 CCGCTCGCGC-----GAGACGCTGCGCGCT 839
DB 279 nArgArgGlyGlyValArgMetSerMetLeuValSerAlaAspAlaLysSerMetProLe 299

```

```

OY 840 GATC-----GAGCG 848
DB 299 uAlaValHisValPheGlyLeuGlyThrArgValLeuAsnLysLeuLeuGlySerGlnVal 319
OY 849 CGGCATGGCGCGCGACGACCTTCGCGGACCGCAT----- 885
DB 319 ArgPheGlnHisLeuProValProPheGlyLeuTyrSerAspGlyTyrLeuAspProVal 339
OY 886 -----CCCTACGTGGCGCGCTCCACCATGCTCAAG-----CAGACCG 923
DB 339 LPheGluGluPheGlyAlaGlyThrAlaAlaLeuHisLeuArgAspGluLeuGluAlaArg 359
OY 924 CGTGTCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945
DB 359 nGluLeuLeuAlaAspArgSerTyrArgArgSerPheArgArgGluPheAspArgIleLeu 379
OY 946 -----ATCATCACTGTGCGCAAGC 965
DB 379 sLeuGlyProSerLeuThrPheHisArgAspPheHisAspAlaValIleValGluCys---Pr 398
OY 966 CTTCGCCGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1025
DB 398 OAspLysSerLeuLeuGlyLysSerPheGlyAlaIleAlaAspGluArgGly----- 415
OY 1026 GTACGACGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1078
DB 416 -----LeuHisProLeuAspAlaPheLeuAspAlaLeuValAspAs 429
OY 1079 ACGAACCCGACGTCGACGCGATCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1138
DB 429 nGlyGluArgAsnValArgThrPheThrIleValAlaAlaHis-----ArgPr 445
OY 1139 TCGCGCGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1195
DB 445 OAsnGluLeuAsnLysLeuAlaAlaGluProSerValHisMetGlyPheSerAspAlaGlu 465
OY 1196 GGCACGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1224
DB 465 yAla-HisLeuArgAsnMetAlaPheTyrAsnPheGlyLeuArgLeuLeuLysArgAla 485
OY 1225 -----CCGCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1252
DB 485 rGAspAlaAspArgAlaGlyGluProPheLeuSerIleGluArgAlaValTyrArgLeu 505
OY 1253 CCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1312
DB 505 hrcIyGluLeuAlaGluTyrPheGlyIle-----GlyAlaGlyThrLeuArgGluGlyAsp 524
OY 1313 TCGCGACCTGTGTGTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1372
DB 524 rGAlaAspPheAlaValIleAspProThrHisLeuAspGluSerValAspGlyTyrHisG 544
OY 1373 CT-----ACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1393
DB 544 IuGluAlaValProTyrTyrGlyGlyLeuArgArgMetValAsnArgAsnAspAlaThrV 564
OY 1394 TCCATTCCTGTACGTCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1452
DB 564 aValAlaThrGlyValGlyGlyThrValValPheArgGlyGlyGlnPheGlyGlyGlnP 584
OY 1453 -----CATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1486
DB 584 heArgAspGlyTyrGlyGlnAsnValLysSerGlyArgTyr----- 597
OY 1487 GAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1530
DB 598 -----LeuArgAlaGlyGluLeuGlyAlaAla 606

```

RESULT 8  
 MUC2\_HUMAN  
 ID MUC2\_HUMAN STANDARD; PRT; 5179 AA.  
 AC 002817; Q14878;



FT CONFLICT 1351 1351 H -> L (IN REF. 3).  
 FT CONFLICT 1412 1412 T -> S (IN REF. 3).  
 FT CONFLICT 1449 1449 L -> P (IN REF. 3).  
 FT CONFLICT 1504 1504 M -> T (IN REF. 3).  
 FT CONFLICT 4192 4192 G -> S (IN REF. 2).  
 SO SEQUENCE 5179 AA: 540295 MW: 85CD7571EB9A5663 CRC64;

## Alignment Scores:

Pred. No.: 8.01e-06 Length: 5179  
 Score: 289.50 Matches: 146  
 Percent Similarity: 32.92% Conservative: 41  
 Best Local Similarity: 25.70% Mismatches: 164  
 Query Match: 8.78% Indels: 217  
 Gaps: 29

US-10-009-782-1 (1-1758) x WUC2\_HUMAN (1-5179)

QY 20 AAGAGAGATTCCATGTCATCCGAT-----CCAGCCCTTCG 61  
 Db 1383 LysilethgvalasnycystprprometaplsycysilethThrProserProPro 1402  
 QY 62 ACCTGCTCTGCGGGCGGACCCCTCATGAGCGGACAGACACCCGGGGGGCGCGCG 121  
 Db 1403 ThrThrThrProserProProProThrThrThrThrThrProPro 1418  
 QY 122 ACCTGGGGGCGGGCGGCGGACGCGCATCGCGGCATCGGAGCGCGCGCGC 181  
 Db 1419 Thr-----ThrThrProserProProThrThrThrThrThrProPro 1434  
 QY 182 ACACCGGGGTCGACGTGCGGGCGCTGTGTGCGCGCGCGCTTCATGACGTGCACACC 241  
 Db 1435 ThrThrThrPro-----SerProProThrThrThrThrThrThrThrPro 1448  
 QY 242 ACAGACAACTACTGCTCTAGGGCGGCGACATGACGCCAACATCGAGGGCGTCA 301  
 Db 1449 LeuProThrThrThr-----Proser 1455  
 QY 302 CCAAGGTCGTCAGGGGCAATGCGGCATACACCTGCGCGCGCGCGACGCCCAACCGC 361  
 Db 1456 ProProThrPro-----ThrThrThrThr 1463  
 QY 362 CCGGCCCCCTGACCTGTCGACGAGAGCGGCTTACGTTTGAGCGCGCTTCGCGACT 421  
 Db 1464 ProProThrThr-----ThrProserProPro 1473  
 QY 422 ACCTGGACGCGTGGCGGCGGCGCGCGCTTCAACGCCGCTGTATGTGGCCATT 481  
 Db 1474 ThrThrThrProserProProThrThrThrProserProProThrThr 1489  
 QY 482 CAACGCTGGCGCGGGGTATCGGACTTGCAGCGCGCGCGCGACGAGAGAAATCG 541  
 Db 1490 -----ThrThrThrThrProProProThrThrThrPro 1500  
 QY 542 CGGCATGCGGGGACCTGCGCGGAGAGCAATGCGCAAGCGCGCGCATCGGATTCACCG 601  
 Db 1501 SerProProThrThrThrProThrPro-----Pro 1511  
 QY 602 GGGCCTTACCGCGCGCGCGCGCGCGCGCACGAGAGATCATGAGTGTCCGCG 661  
 Db 1512 AlaSerThrThrThrLeuProProThrThrThrPro----- 1523  
 QY 662 CGCTGAGCGCGCATGCGGCGCATGCGCACCCACATGCGCGAGAGCGACATCG 721  
 Db 1524 -----SerProProThrThrThrThrThr 1532  
 QY 722 TGGCGCGCGTGAGAGAACTTCGCGATCGCGCGAGCTGAGCGCGGTGATATCT 781  
 Db 1533 ProProProThrThrThrProserProProThr----- 1543  
 QY 782 GCGACACAAAGTATGCGCGCGCGCGCGCGCGCGCGAGAGCGTGCCTGA 841  
 Db 1544 -----ThrThrProThrThr----- 1548

QY 842 TCGAGCCCGCATGGCGGCGCGACGATGCTGAGCGGATCCCT---ACGTCGCG 898  
 Db 1549 -----ProPro-----ThrSerThrThrThrLeuProProThrThrPro 1562  
 QY 899 GCTCCACCATGCTCAACGAGACGCGGTGCTGCGCGGACGACATCATCACTGCT 958  
 Db 1563 SerProProProThrThrThrThr-----ProProProThr 1575  
 QY 959 GCAAGCCCTCCCGCAATGAGCGGGCGGACCTGATGAAATGCGCGGCGAGCGGCA 1018  
 Db 1576 ThrThrProserPro-----ProThrThr 1584  
 QY 1019 AATCCAGT-----ACGAGTGTGTCGCGGACGTCGAGCGGCGGCGCA 1063  
 Db 1585 ThrProserProProThrThrThrThrThrThrProProProThrThrPro 1604  
 QY 1064 TCTACTCATATGACCAAC-----CGACGTGACAGCATCTGCGCTGCGCGCA 1117  
 Db 1605 ProThrThrThrThrThrThrThrProProThrThrThrThrThrProThrThr 1624  
 QY 1118 CCATGATCGCTCGAGCGGCTGCGCGGACGAGCGGCGGCGCATCCGCGCTGGGCA 1177  
 Db 1625 ProThrThrProThrThr-----ThrThr----- 1633  
 QY 1178 CTTCCGCGGGGCTGCGGCACTATGCGGCGGACCTGCGCTGTCGCGGAGACG 1237  
 Db 1634 -----ThrLeuProProThrThr----- 1640  
 QY 1238 CGGTATGAGATGACGGGCTGACCGCGCGCGCGCTTGGCGCGCGGCGGCGACG 1297  
 Db 1641 -----ProserProProProThrThrThrThrThrPro----- 1651  
 QY 1298 TGCAGCGCGGTACTTGCAGACCTGCTGTTCGACCGCGGCGGCGGCGGATA--- 1354  
 Db 1652 ---ProProThrThrThrProserProProThrThrThrThrProserProThr 1670  
 QY 1355 -----CGGCACCTTGAACACCTTACGAGCGCGCG----- 1387  
 Db 1671 ThrThrThrThrProProProThrThrThrProserSerProThrThrThrPro 1690  
 QY 1388 -----CGGCATCCATCCGTGT 1405  
 Db 1691 ProProThrThrThrThrThrThrThrProserProThrThrThrProser 1709  
 QY 1406 ACCTCAACGCGCGCGGCTGTCGCAAGACGAGCGTTCAACGCGGACGAGCGCGCG 1465  
 Db 1710 ThrThrThrThrThrProser-----SerThrThrThrProserProProThrThr 1727  
 QY 1466 TGCTGACAGCAGCGGCGGCTGACCGCGGCGGCGGACCTTACAAATCCGCGTGAACGGG 1525  
 Db 1728 MetThr-----ThrProserProThrThrThrProserProThrThrThrMet 1744  
 QY 1526 CGGCGTGGCGCCCTCCCAACCTGAGCAAAACGGTACATGCGCGCTCCGCGCTC 1585  
 Db 1745 ThrThrLeuProProThrThrThrThrThrThrThrThrThrThrThrPro 1763  
 QY 1586 GCAATACGCGCGCGCGCGGATATCG 1609  
 Db 1764 SerThrThrProProThrThrThrThr 1771

## RESULT 9

VNUA\_PRIVKA STANDARD: PRT: 1733 AA.  
 ID VNUA\_PRIVKA AC P33485.  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DE 01-FEB-1994 (Rel. 28, Last annotation update)  
 OS Probable nuclear antigen.  
 OS Pseudorabies virus (strain Kaplan) (PRV).  
 CC Viruses; dsDNA viruses; no RNA stage; Herpesviridae;  
 CC Alphaherpesvirinae; Varicelloviruses.  
 RN NCBI\_TaxID=33703;  
 RN [1]

[illegible][illegible]



```

OY 249 CAACATCTCTGAGCGGACATGAGCGGCAAGATCTGCGAGGGGCTCACACCGGT 308
DB 141 -----GlyArgHisAlaSerAspArgValGlnAspGlyAlaHisPro 154
OY 309 GGCACGGGCAATGCGGCATGACCGCTGGCGCGGCGGACGCAACCGCGCGCGCC 368
DB 155 ArgArgGlnArgLeuArgGlnProArgHisAlaGly--ArgProArgArgArgGln 173
OY 369 CCGGACCTGCTGAGGAGGAGGCTTACCGCTTCCAGCCCTTCCCGCATCTGGA 428
DB 174 ProProArgArgGlyArgSerArgGlyThrHisArgArgHisLeuArgGlnAlaProArg 193
OY 429 CCGCTGGCGGCGGACGCGCGCGCTGACCGCGCTGATGATGGCGCATTCAC--- 485
DB 194 ProAlaValArgGlyProAspGlnAlaArgGlnAlaArgGlyProArgHisArg 213
OY 486 -----GCTGCGCGC-----CGCGGTCATGCGCGCATTCGACCG 518
DB 214 ArgGluArgHisProProThrAlaArgAspValLeuArgGlyGluProGlyHisGlyAsp 233
OY 519 CGCGCGGACGAGGAGAAATCGCGGCATGCGGACCT-----GCGCGA 563
DB 234 GlyHisHisLeuGlnGlyArgArgGlyArgProArgProGlnGlyArgGlnAlaGlyArg 253
OY 564 GGAAGCCAT---GGCGACGCGCGCGCATGCGCATTC-----GACCGCGCGCTT 608
DB 254 GlyAlaHisProProGlnValAlaArgAlaArgGlyLeuAlaAlaGlyGlnAlaArgGly 273
OY 609 CTACCGCGCGCGCGCGCGCGCACACGAGAGATCATCATGAGTGTGCGCGCGCTGAG 668
DB 274 LeuProGluProArgProGlnGlyValAlaArgHisValHisArgGly----- 288
OY 669 CGCGCATGCGCGCATCTACGCGCACCATGCGCGGAGGAGGAGACACATCGTGGCGCG 728
DB 289 -----GlyArgLeuArgGlyArgValGlyGln 297
OY 729 GCTGAGGAGAACTTCGCGCATCGCGCGGACGAGCTGAGCGCGGTGATTCGACCA 788
DB 298 AlaGly-----ProArgPro---GlnValProGlnAspAlaPro 310
OY 789 CAAAGTCATGCGCGGACCCAAATTCGCGCGCTGCGGAGACGCTGCGGTGATTCGAGGC 848
DB 311 GlnGlyGlu-----AspSerGlu 316
OY 849 CGCGCATGCGCGCGGACGAGCTGCTGCGGAGCGGTATCCCTACGTGCGCGCTCCACAT 908
DB 317 ArgArgGlnThrProProArgProHisSerArg----- 327
OY 909 GCTCAAGCAGGACGCGCTGCTGCGGCGGACGACCATCATCACTGGTGCAGGCCCTT 968
DB 328 -----LysArgArgAspThrGlyAlaHisHisArgHisTrp----- 339
OY 969 CCGCGCACTGAGCGGCGCGGACCTGATGATGCGCGCGGAGCGGCAAAATCCAAATA 1028
DB 340 -----ArgArgArgArgArgArgArgHis 348
OY 1029 CGACGTGTCGCGGACGCTGAGCGCGCGCGCATCTCTCAAGATGAGCAACCGCA 1088
DB 349 ArgGlnGlyAlaLeuProAlaAla-----HisProAspArgArgArgArgGly 365
OY 1089 CCGTGAAGCGCATCTGCGCTGCGCGCGGACCATGATCGGCTCGAGCGCGCTCCGACGA 1148
DB 366 ArgArgAlaHisProAspAla----- 372
OY 1149 CGAGCGCGCGGACGCGCGCTGCGGCGCATCTCCGCGGCTGCGGACATATGCGCG 1208
DB 373 AlaAlaArgAlaSerValPro---AlaHisAlaProAlaHisArgGlyArgLeuArgAla 391
OY 1209 CGA-----CCTGGCGCTGTCGCGCGCTGAGAGCGCGGTATGGAAGAT 1250
DB 392 ArgGlySerThrAlaAlaValAlaProArgProLeuProArg----- 404

OY 1231 GACCGCGCTGACCGCGCGCGCTGCGCGGCGGCGGACGCTGACGCGCGGTA 1310
DB 405 -----GlnHisLeuArgArgAspArgGlyArgGlyArgGlyProHis----- 418
OY 1311 CTGCGCGCATGCTGCTGCTTCGACCGCGGACGCGGTGCG 1349
DB 419 ---HisArgGlyGlyMetGlnArgGlnProHisAlaGly 430

RESULT 11
VNUA_PRIVKA
ID VNUA_PRIVKA STANDARD; PRT: 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE Probable nuclear antigen
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OX Alphaherpesvirinae; Varicelloviruses.
NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91021039; PubMed=2171211;
RX Vitek C., Kozmik Z., Paces V., Schlitt S., Schwytzer M.;
RT "Pseudorabies virus: immediate-early gene overlaps with an oppositely
oriented open reading frame: characterization of their promoter and
enhancer regions.";
RL Virology 179:365-377(1990).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
DR EMBL; M34651; AAA7471.1;
DR PIR; B45344; B45344.
FT DOMAIN 112 117 POLY-THR.
FT DOMAIN 179 1733 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1358 1405 POLY-GLY.
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475B95E2 CRC64;

Alignment Scores:
Pred. No.: 3.92e-05
Score: 274.00 Length: 1733
Percent Similarity: 33.43% Matches: 200
Best Local Similarity: 28.94% Mismatches: 229
Query Match: 8.45% Indels: 231
DB: 1 Gaps: 40

US-10-009-782-1 (1-1758) x VNUA_PRIVKA (1-1733)
OY 1755 CTGCGCGCTGAGGAGGTCGAGCGGACGCGCTGCGGACGAGTCTTGAGAGGT 1666
DB 740 ValGlyGluGlyGlyGlnArgGlnArgGlnAlaAlaAlaAlaGly----- 754
OY 1665 CATTTTCGCGGCGCTGCGCGCGCGG----- 1639
DB 755 ---ValProGluArgAlaArgGlyAlaLeuGlyLeuGlyAlaGluLeuValGly 773
OY 1638 CAGCGCGCGCGGACATCTCTGCGGACGATGCGGTGCGGCGGTATTCGAGCGGA 1579
DB 774 GlnArgValValGlnHisHisAlaHisValLeuGlyValGlyTyrLeuProHisPro 793
OY 1578 GGG-----AGGCGCGCATGATGAGCGGTTCGTCGAG 1549
DB 794 GlyGlyValAlaAlaAlaGluArgGlyAlaAlaAlaArgGlyAspValArg-----Gln 810

```





RN [2]  
 RP SEQUENCE OF 1-181 FROM N.A.  
 RX MEDLINE-84270697; PubMed-6462220;  
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,  
 RT Myers J., Williams C., Ramirez F.;  
 RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary  
 RT conservation of a pattern of introns and exons.";  
 RL Nature 310:337-340(1984).  
 RN [3]  
 RP SEQUENCE OF 162-301.  
 RX TISSUE-SKIN;  
 RA MEDLINE-71038625; PubMed-5529814;  
 RT Click E.M., Bornstein P.;  
 RT "Isolation and characterization of the cyanogen bromide peptides from  
 RT the alpha 1 and alpha 2 chains of human skin collagen.";  
 RL Biochemistry 9:4699-4706(1970).  
 RN [4]  
 RP SEQUENCE OF 263-268.  
 RX TISSUE-SKIN;  
 RA MEDLINE-71001508; PubMed-4319110;  
 RT Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;  
 RT "A comparative study of glycopeptides derived from selected  
 RT vertebrate collagens. A possible role of the carbohydrate in fibril  
 RT formation.";  
 RL J. Biol. Chem. 245:5042-5048(1970).  
 RN [5]  
 RP SEQUENCE OF 425-1464 FROM N.A.  
 RX MEDLINE-84080385; PubMed-6689127;  
 RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,  
 RT Prockop D.J.;  
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the  
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation  
 RT of structures that are conserved during evolution.";  
 RL Biochemistry 22:5213-5223(1983).  
 RN [6]  
 RP SEQUENCE OF 1229-1454 FROM N.A.  
 RX TISSUE-BONE;  
 RA MEDLINE-88124208; PubMed-3340531;  
 RT Maekelaie J.K., Raasina M., Vaita A., Vuorio E.;  
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide  
 RT domain.";  
 RL Nucleic Acids Res. 16:349-349(1988).  
 RN [7]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE-88097389; PubMed-3480516;  
 RA Bornstein P., McKay J., Morishima J.K., Devareyalu S., Gellinas R.E.;  
 RT "Regulatory elements in the first intron contribute to  
 RT transcriptional control of the human alpha 1(I) collagen gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).  
 RN [8]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE-85130970; PubMed-2857713;  
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;  
 RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.  
 RT Promoter structure, Alu repeats, and polymorphic transcripts.";  
 RL J. Biol. Chem. 260:2315-2320(1985).  
 RN [9]  
 RP SEQUENCE OF 1-44 FROM N.A.  
 RX MEDLINE-88033098; PubMed-2822714;  
 RA Nosoum C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,  
 RT de Wet W.J.;  
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)  
 RT collagen gene enhance transcription.";  
 RL J. Biol. Chem. 262:15151-15157(1987).  
 RN [10]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE-91184577; PubMed-2010058;  
 RA Kulvaneni H., Tromp G., Prockop D.J.;  
 RT "Mutations in collagen genes: causes of rare and some common diseases  
 RT in humans.";  
 RL FASEB J. 5:2052-2060(1991).  
 RN [11]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE-97255959; PubMed-9101290;  
 RA Kulvaneni H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 RN [12]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE-91374476; PubMed-1895312;  
 RA Byers P.H., Wallis G.A., Walling M.C.;  
 RT "Osteogenesis imperfecta: translation of mutation to phenotype.";  
 RL J. Med. Genet. 28:433-442(1991).  
 RN [13]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE-97169389; PubMed-9016532;  
 RA Dalgleish R.;  
 RT "The human type I collagen mutation database.";  
 RL Nucleic Acids Res. 25:181-187(1997).  
 RN [14]  
 RP VARIANT OI-II CYS-1166.  
 RX MEDLINE-86287390; PubMed-3016737;  
 RA Cohn D.H., Byers P.H., Steinmann B., Gellinas R.E.;  
 RT "Lethal osteogenesis imperfecta resulting from a single nucleotide  
 RT change in one human pro alpha 1(I) collagen allele.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).  
 RN [15]  
 RP VARIANT OI-II ARG-569.  
 RX MEDLINE-87222295; PubMed-3108247;  
 RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;  
 RT "A point mutation in a type I procollagen gene converts glycine 748  
 RT of the alpha 1 chain to cysteine and destabilizes the triple helix in  
 RT a lethal variant of osteogenesis imperfecta.";  
 RL J. Biol. Chem. 262:14737-14744(1987).  
 RN [16]  
 RP VARIANT OI-II CYS-926.  
 RX MEDLINE-88033031; PubMed-3667599;  
 RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;  
 RT "A point mutation in a type I procollagen gene converts glycine 748  
 RT of the alpha 1 chain to cysteine and destabilizes the triple helix in  
 RT a lethal variant of osteogenesis imperfecta.";  
 RL J. Biol. Chem. 262:14737-14744(1987).  
 RN [17]  
 RP VARIANT OI-II ARG-842.  
 RX MEDLINE-88298828; PubMed-3403550;  
 RA Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;  
 RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)  
 RT chain in lethal perinatal osteogenesis imperfecta. Demonstration of  
 RT the peptide defect by in vitro expression of the mutant cDNA.";  
 RL J. Biol. Chem. 263:11627-11630(1988).  
 RN [18]  
 RP VARIANT OI CYS-1195.  
 RX MEDLINE-89218628; PubMed-3244312;  
 RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;  
 RT "A cysteine for glycine substitution at position 1017 in an alpha  
 RT 1(I) chain of type I collagen in a patient with mild dominantly  
 RT inherited osteogenesis imperfecta.";  
 RL Mol. Biol. Med. 5:197-207(1988).  
 RN [19]  
 RP VARIANT OI-II VAL-434.  
 RX MEDLINE-89255493; PubMed-2470760;  
 RA Patterson E., Smiley E., Bonadio J.;  
 RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta  
 RT mutation.";  
 RL J. Biol. Chem. 264:10083-10087(1989).  
 RN [20]  
 RP VARIANT OI-IV SER-1010.  
 RX MEDLINE-89308591; PubMed-2745420;  
 RA Marini J.C., Grange D.R., Gottesman G.S., Lewis M.B., Koepflin D.A.;  
 RT "Osteogenesis imperfecta type IV. Detection of a point mutation in  
 RT one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";  
 RL J. Biol. Chem. 264:11893-11900(1989).  
 RN [21]  
 RP VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.



```

QY 1289 GCGGCGAGCTCGAGCGCGGCTACTTCCGCGACCTGCTGTGTGACGCCCGGCGAGCTGG 1348
DB 825 Alalysgly---GluProGlyAspAlaGlyala-----LysGly 836
QY 1349 CCGATACCGCCACCTTCGACACCCCTACCGAGCGCGCGGCGATCATTCGCTAGC 1408
DB 837 AspAlaGlyProGlyProGlyProAlaGlyProAlaGlyProProGly-----ProIleGly 854
QY 1409 TCAAGGGGGCGCGCTGCGCAAGAGCGGCGTTCACCGCGGCGGCGGCGGCGGCGGCG 1468
DB 855 AsnValGlyAlaProGlyAlaValGlyAlaValGlySerAlaGlyProProGlyAlaThr 874
QY 1469 TCGCAGCGAGCGCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1516
DB 875 GlyPhe-----ProGlyAlaIleAlaIleValGlyValGlyProProGlyProSer 889
QY 1517 TGAAGGGGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1573
DB 890 -GlyAsnAlaGlyProProGlyProGlyProGlyProAlaGlyLysGlyLysGlyLys 1616
QY 1574 CTCCTCCGCTCGCAATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1616
DB 909 o-----ArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyProProGlyPr 1616
QY 1617 -GGAAGTATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1672
DB 927 oProGlyProAlaGlyGlyLysGlySerProGlyAlaAspGlyProAlaGlyAlaProG 1672
QY 1673 AGAGCT-----GTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1699
DB 947 ythrProGlyProGlyGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLys 1699
QY 1700 CGGCTCGACCGCTGTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1733
DB 967 gGlyGluArgGlyPheProGlyLeuProGlyPro 978

RESULT 13
ID CAEP_RIFPA STANDARD; PRT: 1027 AA.
AC P30734;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE Fibrin-forming collagen alpha chain.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Pogonophora; Vestimentifera; Axonobranchia;
OC Rattidae; Rattidae; Rattidae.
OX NCBI_TaxID=6426;
RN [1]
RP SEQUENCE.
RX MEDLINE=93130909; PubMed=1483468;
RA Mann K., Gall F., Timpl R.;
RT "Amino-acid sequence and cell-adhesion activity of a fibrin-forming
RT collagen from the tube worm Rattus norvegicus living at deep sea
RT hydrothermal vents."
RL Eur. J. Biochem. 210:839-847(1992).
RN [2]
RP SEQUENCE OF 8-45; 525-618 AND 810-882.
RC TISSUE=Cuticle;
RX MEDLINE=92015209; PubMed=1920405;
RA Gall F., Wiedemann H., Mann K., Kuhn K., Timpl R., Engel J.;
RT "Molecular characterization of cuticle and interstitial collagens
RT from worms collected at deep sea hydrothermal vents."
RL J. Mol. Biol. 221:209-223(1991).
CC -1- FUNCTION: FIBRIL-FORMING COLLAGEN.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- PTM: ALL THE "X" ARE PROBABLY GLYCOSYLATED HYDROXYLYSINES.
DR PIR, S22915; S22915.
DR InterPro, IPR000087; Collagen.
DR Pfam, PF01391; Collagen; 16.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen.
FT DOMAIN 1 12 NONHELICAL REGION (N-TERMINAL).
FT MOD_RES 13 1023 TRIPLE-HELICAL REGION.
FT MOD_RES 1024 1027 NONHELICAL REGION (C-TERMINAL).
FT MOD_RES 21 21 HYDROXYLATION (PARTIAL).
FT MOD_RES 24 24 HYDROXYLATION (PARTIAL).
FT MOD_RES 27 27 HYDROXYLATION (PARTIAL).
FT MOD_RES 29 29 HYDROXYLATION.
FT MOD_RES 33 33 HYDROXYLATION.
FT MOD_RES 39 39 HYDROXYLATION.
FT MOD_RES 43 43 HYDROXYLATION.
FT MOD_RES 45 45 HYDROXYLATION.
FT MOD_RES 46 46 HYDROXYLATION.
FT MOD_RES 48 48 HYDROXYLATION.
FT MOD_RES 50 50 HYDROXYLATION.
FT MOD_RES 53 53 HYDROXYLATION.
FT MOD_RES 56 56 HYDROXYLATION.
FT MOD_RES 58 58 HYDROXYLATION.
FT MOD_RES 60 60 HYDROXYLATION.
FT MOD_RES 61 61 HYDROXYLATION.
FT MOD_RES 62 62 HYDROXYLATION.
FT MOD_RES 63 63 HYDROXYLATION.
FT MOD_RES 64 64 HYDROXYLATION.
FT MOD_RES 65 65 HYDROXYLATION.

```



QY 1296 GGTGAGGC-----CGGTACTTCGCGGACCTGCTGTCTGCA----- 1334  
 DB 680 PROARGLYLEUAlaGLYsArGLYLeuArGLYAlaGLYserArGLYgluThr 699  
 QY 1335 -----CCGGCCAGCGGTGGCGGATCCGATCCGTCGTCACAGGCGC 1361  
 DB 700 GLYAlaGLYgluLeuLeuLeuProGLYserProGLYgluProGLY-LeuProGLY 719  
 QY 1362 CTTCGAAACCCCTA---CCGAGCGCGCGCGGATCCATCCGTCGTCACAGGCGC 1418  
 DB 719 roSerGLYgluProGLYProSerGLYPro-----AlaGLYThrAlaG 733  
 QY 1419 GCGGCTGTCGGC-----AAGACAGCGCGTCACGCGGACGATCGCGCGCGCTCCGC 1472  
 DB 733 LyIsGLYglYVal\*\*eLYAlaArGLYserProGLYLeuValGLYsGLYAspa 753  
 QY 1473 AGCAGCGCGCGCTGACGCGCGC-----GCCAGCCCTTACA 1508  
 DB 753 rglYserArspGLYgluProGLYArspGLYThr\*\*gLYgluArGLYgluAspGLY 773  
 QY 1509 ATCGGCGGTACAGCGCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCTACAT 1568  
 DB 773 roProGLYValSerGLYProThrGLYAlaProGLYglu-GlnGLYgluArGLYgluMetPro 792  
 QY 1569 GGC---CCCTCCCTCCGCTGCAATACGCGCGCGCGCGCGCGCGCGCGCGCT 1625  
 DB 793 GLYMetValGLYLeuArGLYgluThrGLYProMetGLYglYgluGLYMet\*\*GLYAsp 812  
 QY 1626 GCG 1685  
 DB 813 GLYGLYProPro-----GLYProSerGLYAsp-----Arg 822  
 QY 1686 GCGCAGCGCGCATCGCGCTGTCACCGCTGTCACAGCGCGCGCGCGCGCGCGCT 1742  
 DB 823 GLYgluArg-----GLYAsnAlaGLYProGlnGLYPro 833

DR InterPro: IP0001007; VWF-C.  
 DR Pfam: PF01391; Collagen; 18.  
 DR Pfam: PF01410; COLF1; 1.  
 DR ProDom: PD000007; Collagen; 1.  
 DR ProDom: PD002078; Fib-collagen; 1.  
 DR SMART: SM00038; COLF1; 1.  
 DR SMART: SM00214; VWF; 1.  
 DR PROSITE: PS01208; VWF; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 222  
 FT PROPEP 23 137  
 FT CHAIN 138 1214  
 FT PROPEP 1215 1460  
 FT DOMAIN 158 174  
 FT DOMAIN 175 1188  
 FT DOMAIN 1189 1214  
 FT SITE 741 743  
 FT SITE 1089 1091  
 FT CARBOHYD 1361 1361  
 SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;  
 Alignment Scores:  
 Pred. No.: 6,76e-05  
 Score: 268.50  
 Percent Similarity: 31.92%  
 Best Local Similarity: 27.70%  
 Query Match: 8.14%  
 DB: 1  
 Gaps: 29  
 Length: 1460  
 Matches: 177  
 Conservative: 27  
 Mismatches: 247  
 Indels: 188

US-10-009-782-1 (1-1758) x CALL\_CANFA (1-1460)

QY 106 CCGGCG 151  
 DB 435 PROGLYsArGLYLeuArGLYAlaGLYserArGLYgluProGLYThrGLYglu 151  
 QY 152 -----CCATCGCGCATCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 185  
 DB 454 YPROGLYProAlaGLY-GlnGLYgluYsArGLYAlaArGLYgluProGLYPro 185  
 QY 186 -----CCGGTTCGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 228  
 DB 474 hGLYLeuProGLYProProGLYgluArGLYglYglYProGLYserArGLY 228  
 QY 229 GATCGCACACCCAGCAGCAGCACTACTGCTGAGCGCTGCGCATGAGCCCAAGATC 288  
 DB 490 ----- 288  
 QY 289 TCGAGGCGGTGACACAGGTGTGTCAGGGGCAATTGCGCATGCGCGCGCGCGCGCG 348  
 DB 491 ----- 348  
 QY 349 CACGCGCAACCG 497  
 DB 497 al-----AlaGLYProGLYglYProAlaGLYgluArGLY 497  
 QY 409 CGCTTCGCGCACTACTGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 507  
 DB 508 -----GlyserProGLYProAlaGLYProLYs----- 516  
 QY 469 ATGTGGCGCATTCACAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528  
 DB 517 --GlyserProGLYgluAlaGLYArGLYProGLYgluAlaGLYLeuProGLYAlaGLY 528  
 QY 529 GACAGGAAGATTCG 536  
 DB 536 eArGLYserProGLYserProGLYPro--AspGLYsThrGLYProProGLYProAl 555  
 QY 589 GCGATTTCAGCG 634  
 DB 555 aGLYgluAspGLYArGLYProGLY-----ProProGLYProProGLYAlaArGLYgl 572

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 DR EMBL: AF153062; A034619.1;  
 DR InterPro: IP000087; Collagen.  
 DR InterPro: IP000085; Fib.collagen.C.



```

QY 635 -----CCGAGAGATCATGAGGTGCGCCGCGCTGAGC 669
DB 572 nAlaGlyValMetGlyPheProGlyProLysGlyAlaAlaGlyLysAlaGly 592
QY 670 GCGCATGGCGCATCTACGCCACCATCGCGAGAGAGAGAGAGAGAGAGAGAG 729
DB 592 yGluArgGlyValProGlyProLysGlyAlaAlaGlyProLysGlyAlaGly 612
QY 730 CTGAGAGAAACCTTCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 783
DB 612 lAlaGlyAlaGlyAlaGlyProGlyProLysGlyAlaAlaGlyProLysGlyAlaGly 631
QY 784 CACCAAGAGATCAT-----GGGCGAGCCCAATTCGCGCGCTG 822
DB 631 lProAlaGlySerProGlyPheGlyGlyLeuProGlyProLysGlyAlaGly 647
QY 823 CCGGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 876
DB 647 rProGlyGlyAlaGlyLysProGlyGlyGlyGlyValProLysPheGlyAlaProG 667
QY 877 GACGCGTATCCCTACGTGCGCGCGCTCCACAT----- 908
DB 667 lProSerGlyAlaArgGlyGlyLysGlyPheProGlyGlyAlaGlyGlyAlaProP 687
QY 909 --GCTCAAGCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966
DB 687 rGlyProAlaGlyProArgGlyAlaAlaGlyAlaProLysAlaAlaGlyAlaGly 706
QY 967 TTCCCGCAAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1026
DB 706 yAspAlaGlyAlaProGlyAlaProGly-----SerGlyGly-----AlaProGlyLeuGly 724
QY 1027 TACGACGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1065
DB 724 lYmetProGlyGlyAlaGlyAlaAlaGlyLeuProGlyProLysGlyAlaProGlyAla 744
QY 1066 TACTTATGATGAGCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1125
DB 744 lAlaGlyProLysGlyAlaAlaGlySerProGlyLysAlaGlyAlaGly-----G 760
QY 1126 GCGTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1185
DB 760 lYleuThrGlyProLysGlyProGlyProLysGlyAlaGlyAlaProGlyAlaGly 780
QY 1186 CCGGT-----GCTGGGCGCATATGCGCGCGCGCGCGCGCGCGCGCGCG 1233
DB 780 lAlaGlyProSerGlyProAlaGlyProThrGlyAlaArgGlyAlaProGlyAlaProGlyG 800
QY 1234 ACGCGGTATGAAAGATGACG-----GCCTGACCGCGCGCGCGCGCGCG 1284
DB 800 lProGlyProProGly-----ProAlaGlyPheAlaGlyProProGlyAlaAlaProGly 819
QY 1285 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1344
DB 819 oGlyAlaGlyGly-----GluProGlyAlaAlaGlyAla-----Lys 831
QY 1345 GTGGCGCATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1404
DB 831 sGlyAlaAlaGlyProProGlyProAlaGlyProThrGlyProProGly-----Pro 849
QY 1405 TACGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1464
DB 849 eGlyAlaAlaGlyAlaGlyProGlyProLysGlyAlaArgGlySerAlaGlyProProGlyAla 869
QY 1465 GTGCTCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1512
DB 869 aThrGlyPhe-----ProGlyAlaAlaGlyAlaGlyAlaGlyProProGly 884
QY 1513 GCGGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1572
DB 884 oSerGlyAlaAlaGlyProProGlyProProGlyProAlaGlyLysGly-----GlyGly 903
QY 1573 CCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1616

```

```

DB 903 yGlyAlaAlaGlyGlyGlyGlyGlyProAlaGlyAlaProGlyGlyValGlyProProGly 923
QY 1617 -----GGAAGTATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1671
DB 923 rProGlyProAlaGlyGlyGlyGlyGlySerProGlyAlaAlaProGlyProAlaProG 943
QY 1672 CAAGACCT-----GTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1698
DB 943 lYThrProGlyProAlaGlyGlyAlaGlyGlyAlaGlyGlyAlaGlyGlyLeuProGlyAla 963
QY 1699 GCGGTCTGACCGCTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1733
DB 963 rGlyGlyAlaGlyPheProGlyLeuProGlyPro 974

RESULT 15
SE5_RAT
ID SE5_RAT STANDARD; PRT; 825 AA.
AC 063003;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SE5 antigen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Brain;
RA MEDLINE=9601519; PubMed=8537300;
RX Suzuki E., Kojima N., Yoshimura K., Dyemura K., Obata K., Akagawa K.;
RT Cloning and sequence analysis of cDNA for a possible DNA-binding
RT protein SE5 in the nervous system."
RL J. Biochem. 118:122-128(1995).
CC - FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - TISSUE SPECIFICITY: EXPRESSED IN NEURONS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
DR EMBL; D37934; BA07153.1;
KW DNA-binding; Nuclear protein; Antigen.
SQ SEQUENCE 825 AA; 86831 MW; AF667FE2FD555BDF CRC64;

Alignment Scores:
Pred. No.: 8,226-05
Score: 267.00 Length: 825
Percent Similarity: 32.33% Matches: 158
Best Local Similarity: 26.33% Conservative: 36
Query Match: 8.09% Mismatches: 221
DB: 1 Indels: 185
Gaps: 28

US-10-009-782-1 (1-1758) x SE5_RAT (1-825)
QY 3 ATTCACCTGATTCGCGGAGAGAGATTCATGTCCTCCATTCGATTCGCGCGCTTGA 62
DB 315 ValProLysAspArgGlyGlu----- 321
QY 63 CCGTGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122
DB 322 ---GlyGlyArgGlyGlyGlyProGlyAlaAlaGlyGlyGlyGlyGlyAlaAlaArg-As 340
QY 123 CCGGAGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
DB 340 P-TyrPheSerGlySer-----ProArgT 348

```

QY 183 CACCCGGGTGACGTGTGGGCGCTGTGGCGCCGCGCTTCATCGACTCGCACACCA 242  
 Db 348 hrLeuGlyGluAspAlaArgAspTrp-----GlySerSerSerAlaAspAla 364  
 QY 243 CGACGACACTACTGCTCAGCGGTGCGGCATGACGCCACAGATCTCGAGCGCTGCAC 302  
 Db 364 laGlySerSerProCysAlaLeuArgGlySerLeuAlaProGluArgLeuGlyAspGlyP 384  
 QY 303 CACGCTGTGTCAGCGCATTTGCGGCATCAGCTGGCGCGCTGCCACGCCAACCCGCC 362  
 Db 384 ro-----TrpAlaTrpProSerProGluGlu 394  
 QY 363 CGCCCCCTGGACCTGCTGACGAGCGCGCTTACCGTTTCAGACGCTTCGCGACTA 422  
 Db 394 rgluPro----- 396  
 QY 423 CCTGACGCGTTGCGGGCACCGCGCGCGCTCAACGCCGCTGTATGTGGCCATTC 482  
 Db 397 -----GlyProAlaArgValGluSerProArgLutTrpGly----- 409  
 QY 483 AACGTCGCGCGCGGTATGCGGACTTTCAGCGCGCGCGCACCGAGAGAAATGCG 542  
 Db 409 ----- 409  
 QY 543 GGCATGCGGACCTGGCGGAGAGAACATGCGCACGCCGCTATGCGCATTTGCACCGG 602  
 Db 410 -----GlyThrGluSerProAlaGlyTrpGluAlaGlyProArgLutTrpGlyPro 427  
 QY 603 CCGCTTCTACCGCGCGCGCGCGCGCGCACCGAGAGATCATCGAGGTGTGCGCGCC 662  
 Db 427 exProGlyGlyArgGlyAsp-----GlyProAlaArgArgProAlaArgArgArgGly 445  
 QY 663 GCTGACGCGCATGGCGGCATCTACGCCACCATCGCGCGCGAGAGAGCGAGCATCGT 722  
 Db 445 rgluGlyArgMetGlyArgGluLeuGluThrAlaThrSerAlaSerAlaThrGlyG 465  
 QY 723 GCGCGCGCTGAGGAACTTCGCGATCGCGCGCGAGGTGACCTGCGCGGTGTGATCTC 782  
 Db 465 lProAla-----GluGluAlaGlyAlaSerAlaProGlu-GlyGluAlaGlyGlyPro 483  
 QY 783 GCACCAACAAGTCAATGGGCCAGCCCA-----TTTCGCGCGCTGCGCGCA 827  
 Db 484 ArgGlyArgAlaArgGlyProArgGluAlaAlaArgArgArgHisGlyProGluArg 503  
 QY 828 GACGCTGCGCTGATCGAGCGCGCATGGCGCGCGAGAGCTGCTCGT----- 875  
 Db 504 ArgGlyProProGluAlaGlyGluGly---ProGlyAspAlaThrLeuValLeuGly 522  
 QY 876 ---GGACGCTATCCCTACGTGGCGCGCTCCACCATGCTCAAGCAGAGCGCTGCTGT 932  
 Db 523 LeuGlyThrThrSerGlyGluGluAlaAspGlnSerGlnThrLeuProAlaLeuAla 542  
 QY 933 GGCAGG---ACGCACATCATCACTGTGTGCAAGCCCTTCCGAACTGAGCGCGCGCA 989  
 Db 543 GluAlaProThrAlaHisAlaHisAlaValProGly---ProGlyProAlaAlaAlaThr 561  
 QY 990 CTTGATGAAGTCGCGCGCGCGCGCAATCAAGTAGAGAGCTGTGCGCGCA----- 1043  
 Db 562 LeuGly-----GlyArgGlyArgArgGlySerTrpArgGlyGlyArgArgGlyGly 578  
 QY 1044 -----GTCGACACCGCGCGCGCATCTTACATGATGAGAGCA-----ACCGCA 1088  
 Db 579 GluAlaGlyAlaSerGlyGlyArgGlyGlyArgGlyArgGlyArgGlyArgGlyArg 598  
 QY 1089 CGTGACGACATCTGCGCTGCGCGCGCATGATCGGCTCCGAGCGCTGCGCGCACGA 1148  
 Db 599 GlySerGlyLeuSerGlyThrArgGluAsp-----AlaGlySerProSerAlaArg 615  
 QY 1149 CGAGCGCGCGCATCGCGCGCTGTGGGCACTTCCCGCGGCTGTGGGCACTATGC--- 1205  
 Db 616 ArgGlyGluGluArgArgArg---GlyHisGlyProProAlaAlaGlyAlaAlaGlnVal 634  
 QY 1206 -----GCGCGACCTGGGCGCT 1220

Db 635 SerThrArgGlyArgArgAlaArgGlyGluArgGlyGlyGluGluAlaAspGlyLeu 654  
 QY 1221 GTTCCCGCTGGAGAGCGCGGTATGGAAGATGACCGGCGCTGACCGCGCGCTTGCGCT 1280  
 Db 655 LeuProAlaGly-----ArgAspArgLeuProLeuArgPro 666  
 QY 1281 GCGCGCGCGCGGAGCTGACGAGCGCGGACTTCCCGACCTGCTGCTGTGTTGACCGCGC 1340  
 Db 667 Gly-----AspSerAsnGlnArgValGluArgProGlyHisProArgGlyGly 682  
 QY 1341 CACGCTGCGCATACCGCGCACCTTGAACACCTTACCGAGCGCGCGCGCATGCAATC 1400  
 Db 683 HisGlyAla-----IleAsnAlaProSerAlaProAspAlaSerProProHisAlaProArg 701  
 QY 1401 CGT---GTACGTCAACGCGCGCGCGGTGTGCGAAGAGCAGCGCTTACCGCGCGCATGCG 1457  
 Db 702 ArgTrpValSerGlnArgGlnArgGluArgLeuTrpArgGln----- 714  
 QY 1458 CGCGCGCGTCTGCGACGCGACGCGCGCGCTGAGCGCGCGCGCATACATCGGCGCT 1517  
 Db 715 -----PheArg 716  
 QY 1518 GAAGCGCGCGCGCTGCGCGCGCGCTTCCCAACCTTGGAGCAACCGCTACATGCGCGCTCC 1577  
 Db 717 ValGlyGlyGlyPheProProProProProProProProProProProProProPro 736  
 QY 1578 CTCGCG-----TCGCAATACGCGCGCGCGCATGCTGCGCGCATGAGAGTAT 1625  
 Db 737 LeuArgLeuThrCysAlaGlyAspProGlyAlaSerArg-----ProGlySer--- 752  
 QY 1626 GCGCGCGCGCGCTGCGCGCGCGCGAGCGCAAGCGCGGAAATGACCTGCAAGACT 1679  
 Db 753 ---ArgArgProAlaArgArgProArgGly---GluLeuThrProGlnArgPro 768

Search completed: May 11, 2003, 12:10:53  
 Job time : 73 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 01:04:21 ; Search time 4570 Seconds

(without alignments)  
1195.346 Million cell updates/sec

Title: US-10-009-782-1

Sequence: 1 gaattccactgacgagcga.....cccgagctacgagagcct 1758

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank :\*

1: gb\_ba:\*

2: gb\_hg:\*

3: gb\_in:\*

4: gb\_ov:\*

5: gb\_ov:\*

6: gb\_ov:\*

7: gb\_ov:\*

8: gb\_ov:\*

9: gb\_ov:\*

10: gb\_ov:\*

11: gb\_ov:\*

12: gb\_ov:\*

13: gb\_ov:\*

14: gb\_ov:\*

15: gb\_ov:\*

16: gb\_ov:\*

17: gb\_ov:\*

18: gb\_ov:\*

19: gb\_ov:\*

20: gb\_ov:\*

21: gb\_ov:\*

22: gb\_ov:\*

23: gb\_ov:\*

24: gb\_ov:\*

25: gb\_ov:\*

26: gb\_ov:\*

27: gb\_ov:\*

28: gb\_ov:\*

29: gb\_ov:\*

30: gb\_ov:\*

31: gb\_ov:\*

32: gb\_ov:\*

33: gb\_ov:\*

34: gb\_ov:\*

35: gb\_ov:\*

36: gb\_ov:\*

37: gb\_ov:\*

38: gb\_ov:\*

39: gb\_ov:\*

40: gb\_ov:\*

41: gb\_ov:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1758	100.0	1758	BD011857	BD011857 Transform
2	1758	100.0	1758	E59056	E59056 Transform
3	1758	100.0	1758	BD004254	BD004254 Transform
4	1739.4	98.9	1980	S80683	S80683 dan-D-amlino
5	1433.4	82.7	1455	AFANADA	AF332548 Alcaligenes
6	1088.6	61.9	1455	AF332548	AF332548 Alcaligenes
7	693.6	39.5	1497	AFPA61	D45919 Alcaligenes
8	533.6	30.4	208050	AL646083	AL646083 Ralstonia
9	455.8	25.9	1467	AFPA6	D50061 Alcaligenes
10	306	17.4	110000	LMFLCHR34_10	Continuation (11 O
11	171.8	9.8	22251	SC2K36	Continuation (11 O
12	116	6.6	82746	AF453501	AF453501 Actinomyces
13	107.4	6.1	28751	SC6A11	AF453501 Actinomyces
14	99.2	5.6	14710	AE006004	AE006004 Caulobact
15	99.2	5.6	123580	AX211705	AX211705 Sequence
16	99.2	5.6	123580	AF263912	AF263912 Streptomy
17	99.2	5.6	123580	AX211739	AX211739 Sequence
18	97.2	5.5	113193	AF357202	AF357202 Streptomy
19	94.2	5.4	5123	HEPVE	AF357202 Streptomy
20	94.2	5.4	8438	SHILT	AF357202 Streptomy
21	94	5.3	4039	AF425991	M5120 Pseudorabid
22	94	5.3	34766	SCC121	M5120 Pseudorabid
23	93.6	5.3	110000	LMFLCHR32_06	AF425991 Streptomy
24	93.4	5.3	27541	SC6D7	AF425991 Streptomy
25	91.4	5.2	37501	AX211706	Continuation (7 of
26	91	5.2	35033	SC6D7	AX211706 Sequence
27	91	5.2	35654	SC7H1	AL133213 Streptomy
28	89.8	5.1	15141	SHIPROIE	AL096811 Streptomy
29	89.6	5.1	33779	SCGD3	AL021411 Streptomy
30	88.8	5.1	110000	LMFLCHR36_31	M3451 Pseudorabid
31	88.2	5.0	138404	SC2G5	AL096822 Streptomy
32	88	5.0	138404	AE005104	Continuation (32 O
33	87.8	5.0	1746	AF425992	AL035478 Streptomy
34	87.8	5.0	30657	SCD12A	AE005104 Halobacte
35	86.8	4.9	41906	SC5C7	AF425992 Streptomy
36	85.6	4.9	88421	AX417445	AL357524 Streptomy
37	85.6	4.9	5520	AC084329	AL035151 Streptomy
38	85.6	4.9	49736	AF319998	AX417445 Sequence
39	85.2	4.8	25362	SCF56	AC084329 Leishmani
40	84.6	4.8	109519	AX195929	AF319998 Stigmatal
41	84	4.8	1839	AX196078	AX195929 Sequence
42	83.8	4.8	3578	AF005277	AX196078 Sequence
43	83.8	4.8	3619	SCVALSFP	AF005277 Cellulose
44	83.8	4.8	24800	SC9B1	Y13070 S.coelicolo
45	83.8	4.8	110000	LMFLCHR36_03	AL049727 Streptomy

## ALIGNMENTS

RESULT 1

BD011857

LOCUS BD011857 1758 bp DNA linear PAT 02-AUG-2002

DEFINITION Transformant microorganism and process for producing

ACCESION BD011857

VERSION BD011857.1 GI:22092046

KEYWORDS WO 0078926-A/1.

SOURCE unidentified.

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 1758)

AUTHORS Takeuchi, K., Koide, Y., Hirose, Y., Moriyuchi, M. and Isobe, K.

TITLE Transformant microorganism and process for producing D-aminocyclase

JOURNAL Patent: WO 0078926-A 1 28-DEC-2000;

Pred. No. is the number of results predicted by chance to have a

## COMMENT

AMANO PHARMACEUTICAL CO LTD, KENICHI TAKEUCHI, YOSHINAO KOIDE,  
YOSHIIKO HIROSE, MITSUAKI MORIGUCHI, KIMIYASU ISOBE  
OS  
ALCALIGENES xylosoxydans subsp. xylosoxydans  
PN  
WO 0078926-A/1  
PD  
28-DEC-2000  
PF  
15-JUN-2000 WO 2000JP003932  
PR  
17-JUN-1999 JP 99P 170555  
PI  
KENICHI TAKEUCHI, YOSHINAO KOIDE, YOSHIIKO HIROSE, PI  
MITSUAKI MORIGUCHI,  
KIMIYASU ISOBE  
PC  
C12N1/21, C12N15/52, C12N9/80  
CC  
FH

## FEATURES

## SOURCE

Key Location/Qualifiers.  
1. 1758  
/organism="unidentified"  
/db\_xref="taxon:32644"

## BASE COUNT

281 a 662 c 568 g 247 t

## ORIGIN

Query Match 100.0%; Score 1758; DB 6; Length 1758;  
Best Local Similarity 100.0%; Pred. No. 9e-183;  
Matches 1758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCACCTTATGCGGAGAGATTTCCATGTCCTCCATCCGATCCAGCCCTTC 60  
DB 1 GAATTCACCTTATGCGGAGAGATTTCCATGTCCTCCATCCGATCCAGCCCTTC 60  
QY 61 GACCTGCTCTGCGGGGCGACCTTCATCGAGGAGACAAACCCCGGGCGCGCC 120  
DB 61 GACCTGCTCTGCGGGGCGACCTTCATCGAGGAGACAAACCCCGGGCGCGCC 120  
QY 121 GACCTGCGGTGCGGGGAGACCGCATCGCCGCATCGCGATCTGTGCGAGCGCGCG 180  
DB 121 GACCTGCGGTGCGGGGAGACCGCATCGCCGCATCGCGATCTGTGCGAGCGCGCG 180  
QY 181 CACACCGGGGTGCGAGTGTGCGGCTGTGTGCGGCGGCTTCATCGACTGCGACAC 240  
DB 181 CACACCGGGGTGCGAGTGTGCGGCTGTGTGCGGCGGCTTCATCGACTGCGACAC 240  
QY 241 CACGACGACAACTACCTGCTCAGGCGCTGCGGACATGAGCGCCAAAGATCTGCGAGGGGCTC 300  
DB 241 CACGACGACAACTACCTGCTCAGGCGCTGCGGACATGAGCGCCAAAGATCTGCGAGGGGCTC 300  
QY 301 ACCACGGTGTACAGGGCAATTCGGGATCAGCTGCGCGCTGCGCGACGCCAACCCG 360  
DB 301 ACCACGGTGTACAGGGCAATTCGGGATCAGCTGCGCGCTGCGCGACGCCAACCCG 360  
QY 361 CCGGCCCCCTGACCTGCTGAGAGAGCGGCTCTTACCGTTTGAAGGCTTCGCGAC 420  
DB 361 CCGGCCCCCTGACCTGCTGAGAGAGCGGCTCTTACCGTTTGAAGGCTTCGCGAC 420  
QY 421 TACCTGAGCGCTTGGGGGCGACGCGCGCGGCTGTAACGCGCGCTGATGATGGGCAT 480  
DB 421 TACCTGAGCGCTTGGGGGCGACGCGCGCGGCTGTAACGCGCGCTGATGATGGGCAT 480  
QY 481 TCAAGCTGCGCGCGCGGTATGCGGACTTTCAGGCGCGCGCGCATCGGCAATTCGACC 480  
DB 481 TCAAGCTGCGCGCGCGGTATGCGGACTTTCAGGCGCGCGCGCATCGGCAATTCGACC 480  
QY 541 GCGGCGATGCGGAGCTGCGCGGAGAGACCATGCGCAGCGCGCGCATCGGCAATTCGACC 600  
DB 541 GCGGCGATGCGGAGCTGCGCGGAGAGACCATGCGCAGCGCGCGCATCGGCAATTCGACC 600  
QY 601 GCGGCGCTTTCACCG 660  
DB 601 GCGGCGCTTTCACCG 660  
QY 661 CCGGTGAGCGGAGTGGCGGCGATCTAGCGCACCATCATGCGGAGAGAGCGCGCGCGCG 720  
DB 661 CCGGTGAGCGGAGTGGCGGCGATCTAGCGCACCATCATGCGGAGAGAGCGCGCGCGCG 720  
QY 721 GTGGCGCGCTGAGGAAACCTTCGCGATCGCGCGGACGTGCGGCTGTGATC 780

DB 721 GTGGCGCGCTGAGGAAACCTTCGCGATCGCGCGGACGTGCGGCTGTGATC 780  
QY 781 TCGACACCAAGATCATGCGGCGACGCCAATTCGCGCGCTGCGGCGAGAGCGCTGCGCTG 840  
DB 781 TCGACACCAAGATCATGCGGCGACGCCAATTCGCGCGCTGCGGCGAGAGCGCTGCGCTG 840  
QY 841 ATGAGGCGCGCATGCGGCGCGAGAGCTGCTGCGTGAAGCGGTATCCCTACGTGCGCGG 900  
DB 841 ATGAGGCGCGCATGCGGCGCGAGAGCTGCTGCGTGAAGCGGTATCCCTACGTGCGCGG 900  
QY 901 TCCACCATGCTCAAGCAGAGACCGCTGCTGCTGCGGCGAGCGACCATCATACCTGCTGC 960  
DB 901 TCCACCATGCTCAAGCAGAGACCGCTGCTGCTGCGGCGAGCGACCATCATACCTGCTGC 960  
QY 961 AAGCCCTTCCCGAAGTGAAGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
DB 961 AAGCCCTTCCCGAAGTGAAGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
QY 1021 TCCAGTACGAGTGTGCGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
DB 1021 TCCAGTACGAGTGTGCGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
QY 1081 GAACCCGAGTGTGCGAGCGCATCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
DB 1081 GAACCCGAGTGTGCGAGCGCATCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
QY 1141 CCGGACGAGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
DB 1141 CCGGACGAGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
QY 1201 TATGCGGCGGACCTGCGGCTGCTCCGCTGAGAGCGGCTGATGAGAGATGAGCGGCTG 1260  
DB 1201 TATGCGGCGGACCTGCGGCTGCTCCGCTGAGAGCGGCTGATGAGAGATGAGCGGCTG 1260  
QY 1261 ACCGCGCGGCTTGGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
DB 1261 ACCGCGCGGCTTGGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
QY 1321 CTGGGTGTTTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
DB 1321 CTGGGTGTTTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
QY 1381 CCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
DB 1381 CCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
QY 1441 TTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
DB 1441 TTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
QY 1501 CCGTACCAATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
DB 1501 CCGTACCAATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
QY 1561 CGCTACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
DB 1561 CGCTACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
QY 1621 GTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680  
DB 1621 GTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680  
QY 1681 TCGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740  
DB 1681 TCGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740  
QY 1741 CTGAGCTACGAGAACTT 1758  
DB 1741 CTGAGCTACGAGAACTT 1758

RESULT 2

[illegible]

D	61	GACCTCTGCTCGCGGGGACCCCTCATGAGGAGCAACACCCCGGGGGCGCGCC	120
Q	121	GACCTGGGGGTGGGGGACCGCATCCCGCGCATGGGATCTGTGGACGCCCGCG	180
D	121	GACCTGGGGGTGGGGGACCGCATCCCGCGCATGGGATCTGTGGACGCCCGCG	180
Q	181	CACACCCGGGTGACGTGTGGGGCTGGTGGCGCCGGGGTTATATGCACTCCGACAC	240
D	181	CACACCCGGGTGAGGTGTGGGGCTGGTGGCGCCGGGGTTATATGCACTCCGACAC	240
Q	241	CACGAGACAACTACCTGCTCAAGCGGTGCGACATGACGCCCAAGATCTCGAGGGCTC	300
D	241	CACGAGACAACTACCTGCTCAAGCGGTGCGACATGACGCCCAAGATCTCGAGGGCTC	300
Q	301	ACCAGGTGTGACAGGGGCAATTGGGGCAATGACCTTGGGGCGCGTGGCGACGCCAACCG	360
D	301	ACCAGGTGTGACAGGGGCAATTGGGGCAATGACCTTGGGGCGCGTGGCGACGCCAACCG	360
Q	361	CCCGCCCCCTGGACCTCTGACACGAGGGGCTTACACCTTTCGAGCGCTTGGCGAC	420
D	361	CCCGCCCCCTGGACCTCTGACACGAGGGGCTTACACCTTTCGAGCGCTTGGCGAC	420
Q	421	TACCTGGAGCGGTTTCGGGGCCACGCGCGCGCTCAACGCCGCTGTATGATGGCCAT	480
D	421	TACCTGGAGCGGTTTCGGGGCCACGCGCGCGCTCAACGCCGCTGTATGATGGCCAT	480
Q	481	TCAACGCTGCGCGCGGTATGCGGACCTTGACGCGCGCGCGCACCGCAGAGAAATC	540
D	481	TCAACGCTGCGCGCGGTATGCGGACCTTGACGCGCGCGCGCACCGCAGAGAAATC	540
Q	541	GCGGCATTCGCGGACCTTGGCGAGAACCATGCGCAGCGCGCATTCGATTCGACC	600
D	541	GCGGCATTCGCGGACCTTGGCGAGAACCATGCGCAGCGCGCATTCGATTCGACC	600
Q	601	GGCGCTTCTTCCCGCGCGCGCGCGCGCCACCAAGAGATCATGAGGTGTCCGG	660
D	601	GGCGCTTCTTCCCGCGCGCGCGCGCGCCACCAAGAGATCATGAGGTGTCCGG	660
Q	661	CCGCTGAGCGCGCATGCGCGCATCTACGCCACCCATGCGCGCAGAGAGCGACATC	720
D	661	CCGCTGAGCGCGCATGCGCGCATCTACGCCACCCATGCGCGCAGAGAGCGACATC	720
Q	721	GTCGCGCGGTGGGAGGAACTTCCCGATGCGCGCGCAGCTGGACGTGCCGGTGTATC	780
D	721	GTCGCGCGGTGGGAGGAACTTCCCGATGCGCGCGCAGCTGGACGTGCCGGTGTATC	780
Q	781	TGCGACACCAAGTTCATGGGCGCAGCCCAATTTGGCGCGCTCGCGCGAGAGGTCGCGTG	840
D	781	TGCGACACCAAGTTCATGGGCGCAGCCCAATTTGGCGCGCTCGCGCGAGAGGTCGCGTG	840
Q	841	ATGAGAGCGCCCATGCGCGCGCGCAGAGGTCCTCGCTGAGCGCTATCCCTACGTGGCGCG	900
D	841	ATGAGAGCGCCCATGCGCGCGCGCAGAGGTCCTCGCTGAGCGCTATCCCTACGTGGCGCG	900
Q	901	TTCACATGCTCAAGAGAGACCGCGGTGTCTGTGGCGACCACTATCACCCTGTGTC	960
D	901	TTCACATGCTCAAGAGAGACCGCGGTGTCTGTGGCGACCACTATCACCCTGTGTC	960
Q	961	AAGCCCTTCCCGAATGAGCGGCGCGACCTTGAGTGAAGTGGCGCGCGCGCAAA	1020
D	961	AAGCCCTTCCCGAATGAGCGGCGCGACCTTGAGTGAAGTGGCGCGCGCGCAAA	1020
Q	1021	TCCAGTACGAGTGTGTCGCGAGCTGACGCCGCGCGCGCATCTACTTATGATGAC	1080
D	1021	TCCAGTACGAGTGTGTCGCGAGCTGACGCCGCGCGCGCATCTACTTATGATGAC	1080
Q	1081	GACCCGAGCTGACGAGCTCTTGGCGCTTGGCGCGCGACCATATGCGGTCCGACGCGCTG	1140
D	1081	GACCCGAGCTGACGAGCTCTTGGCGCTTGGCGCGCGACCATATGCGGTCCGACGCGCTG	1140
Q	1141	CCGACGAGAGCGCGCGCATCCGCGCTGTGGGCGACCTTCCGCGGGTCTGGGGAC	1200



Db 1141 CCGCAGACGAGCGCCGCGATCCGCGCTGTGGGCACTTCCCGGGGTCTGCTGGGCGAC 1200  
 Qy 1201 TATGGCGCGGACCTGGGCTGTTCCTCCGCGAGAGCGCGGTATGAAAGATGACCGGCTG 1260  
 Db 1201 TATGGCGCGGACCTGGGCTGTTCCTCCGCGAGAGCGCGGTATGAAAGATGACCGGCTG 1260  
 Qy 1261 ACCGCGCGCGCTTGGCGCTGCGCGCGCGCGAGCTCAGACCGGGGACTTCCCGGAC 1320  
 Db 1261 ACCGCGCGCGCTTGGCGCTGCGCGCGCGAGCTCAGACCGGGGACTTCCCGGAC 1320  
 Qy 1321 CTGGTGGGTTCGACCGCGCGCGCGCGCGATACCGCGACCTTGAACACCTTACCGAG 1380  
 Db 1321 CTGGTGGGTTCGACCGCGCGCGCGCGCGATACCGCGACCTTGAACACCTTACCGAG 1380  
 Qy 1381 CGCGCGCGCGCGATTCCTGCTGATGCAAGCGCGCGCGCGGTTCGCAAGAGAGCGCG 1440  
 Db 1381 CGCGCGCGCGCGATTCCTGCTGATGCAAGCGCGCGCGGTTCGCAAGAGAGCGCG 1440  
 Qy 1441 TTCACCGCGCGAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
 Db 1441 TTCACCGCGCGAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
 Qy 1501 CCTTACATCCGCGCGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560  
 Db 1501 CCTTACATCCGCGCGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560  
 Qy 1561 CGTACATGCG 1620  
 Db 1561 CGTACATGCG 1620  
 Qy 1621 GTGATGGCG 1680  
 Db 1621 GTGATGGCG 1680  
 Qy 1681 TCGCAGCGCGAGCG 1740  
 Db 1681 TCGCAGCGCGAGCG 1740  
 Qy 1741 CTGAGCTACGAGAGCTT 1758  
 Db 1741 CTGAGCTACGAGAGCTT 1758

RESULT 4  
 S80683 1980 bp DNA linear BCT 13-AUG-1996  
 LOCUS dan-D-aminocyclase [Alcaligenes xylosoxydans, ssp. xylosoxydans,  
 A-6, Genomic, 1980 nt].  
 ACCESSION S80683  
 VERSION S80683.1 GI:1488356  
 KEYWORDS  
 SOURCE Achromobacter xylosoxydans A-6 ssp. xylosoxydans.  
 ORGANISM Achromobacter xylosoxydans  
 Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 Achromobacter.  
 TITLE Y' (bases 1 to 1980)  
 AUTHORS Wakayama, M., Katsuno, Y., Hayashi, S., Miyamoto, Y., Sakai, K. and  
 Morituguchi, M.  
 JOURNAL Cloning and sequencing of a gene encoding D-aminocyclase from  
 MEDLINE Alcaligenes xylosoxydans subsp. xylosoxydans A-6 and expression of  
 PUBMED the gene in Escherichia coli  
 85100942 Biotech. J. Biotechnol. 59 (11), 2115-2119 (1995)  
 REMARK GenBank staff at the National Library of Medicine created this  
 entry [NCBI gidsq 174864] from the original journal article.  
 This sequence comes from Fig. 3.  
 FEATURES  
 source 1..1980  
 Location/Qualifiers  
 /organism="Achromobacter xylosoxydans"  
 /db\_xref="taxon:85698"  
 201..1655  
 /gene="dan"

Query Match 98.9%; Score 1739.4; DB 1; Length 1980;  
 Best Local Similarity 99.4%; Pred. No. 9, 1e-181;  
 Matches 1748; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

BASE COUNT 298 a 758 c 642 g 282 t

ORIGIN

1 GAATTCACCTGATGCGGAGAGAGAGATTCCTGCTCCCAATCGATTCGAGCCCTTC 60  
 Db 168 GATGCGCACTGATGATTCCTCCCGCGGAGATTCCTGCTCCCAATCGATTCGAGCCCTTC 227  
 Qy 61 GACCTGCTGCTCG 120  
 Db 228 GACCTGCTGCTCG 287  
 Qy 121 GACCTGCGGTGCG 180  
 Db 288 GACCTGCGGTGCG 347  
 Qy 181 GACACCGGGGTGCG 240  
 Db 348 GACACCGGGGTGCG 407  
 Qy 241 GACGAGCACTACCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
 Db 408 GACGAGCACTACCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 467  
 Qy 301 ACCACGGTGTACG 360  
 Db 468 ACCACGGTGTACG 527  
 Qy 361 CCCGCCCG 420  
 Db 528 CCCGCCCG 587  
 Qy 421 TACCTGAGCGGTGCG 480  
 Db 588 TACCTGAGCGGTGCG 647  
 Qy 481 TCAACGGTGTGCG 540  
 Db 648 TCAACGGTGTGCG 707  
 Qy 541 GCGGCG 600  
 Db 708 GCGGCG 767  
 Qy 601 GCGGCG 660  
 Db 768 GCGGCG 827  
 Qy 661 CCGCTGAGCG 720

/note="D-aminocyclase, N-acetyl-D-amino acid amidohydrolase"  
 201..1655  
 /gene="dan"  
 /note="3.5.1.14; Method: conceptual translation with  
 partial peptide sequencing; This sequence comes from Fig.  
 3; N-acetyl-D-amino acid amidohydrolase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="D-aminocyclase"  
 /protein\_id="AAB35881.1"  
 /db\_xref="GI:1488357"  
 /translation="MSQSDSPEDPDLIAGGLTIDSSNTPGRRADLGYRDRRIAATGDL  
 SDAATRVYVSGLYVAPGFIQSHTDNDYLLRRDMPTKISQGVTTVTGCGTSLA  
 FLAANPAPVLDLDEGSGYRPERFADYLDALRIAPAAVNAACVGHSTLRANVPLD  
 ORATDEPILAMDLAEEMASGALISITGATVPAPAAATTEELIEVCRPLSANGIY  
 ATMRBDEGHTVALEETRIEGLDLPVYIHHKVMGPNFGRSELTPLLEAMAR  
 OVDYSDVAPYPMASSTMLKODRYLAFRTIITCKPPELSGDLDEVAERKSTYD  
 VPELQPAVGIYEMDEPVDQRIAFPTGPIAGRGOLGQYRADLVPEPDAVADTATPEFTER  
 DLGLPFLTAIVVMKTLTAARFGLAGRGOLGQYRADLVPEPDAVADTATPEFTER  
 AAGIHSVYVNGAPVYVMEQAFQGHGRVLAARTAA"

RESULT 5

AFAANDA  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE

ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

TITLE

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

FEATURES  
source

CDS

BASE COUNT  
ORIGIN

Query Match  
Best Local Similarity 99.98; Pident. No. 1.3e-149;  
Matches 1454; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

34 ATGTCCCAATCGATTCAGCCCTTGACGCGTGGCTGCCGGGGGCACCCTCATGCAC 93  
db 1 ATGTCCCAATCGATTCAGCCCTTGACGCGTGGCTGCCGGGGGCACCCTCATGCAC 60

94 GGACAGCATTACATCC

Db 61 GGCAGCAACACCCCGGGGCGCGCGAGACTGTGGCGTGGCGGCGACCGCATCGCCGCC 120  
 Oy 154 ATCGGCGATCTGTGGAGCGCGCGCGCGACACCGGGGTGAGCTGTGGGCGCTGTGGTTC 213  
 Db 121 ATCGGCGATCTGTGGAGCGCGCGCGCGACACCGGGGTGAGCTGTGGGCGCTGTGGTTC 180  
 Oy 214 GCGCCCGGCTTATGAGTGTGAGACCGCGAGACGACGACGACGACGACGACGACGACGAC 273  
 Db 181 GCGCCCGGCTTATGAGTGTGAGACCGCGAGACGACGACGACGACGACGACGACGACGAC 240  
 Oy 274 ATGAGCGCCCAATATCTCGAGAGGGGCTGACGACGAGTGTGACGGGCAATTTGCGGATCAC 333  
 Db 241 ATGAGCGCCCAATATCTCGAGAGGGGCTGACGACGAGTGTGACGGGCAATTTGCGGATCAC 300  
 Oy 334 CTGGGCGCGCTGTGGCG 393  
 Db 301 CTGGGCGCGCTGTGGCG 360  
 Oy 394 TCTTACCGCTTGTGAGCGCTTGTGCGGAGTGTGAGCGGCTTGTGCGGCGCGCGCGCGCG 453  
 Db 361 TCTTACCGCTTGTGAGCGCTTGTGCGGAGTGTGAGCGGCTTGTGCGGCGCGCGCGCGCG 420  
 Oy 454 GTCAACGCGCGCTGTGAGCGCTTGTGCGGAGTGTGAGCGGCTTGTGCGGCGCGCGCGCG 513  
 Db 421 GTCAACGCGCGCTGTGAGCGCTTGTGCGGAGTGTGAGCGGCTTGTGCGGCGCGCGCGCG 480  
 Oy 514 CAGCG 573  
 Db 481 CAGCG 540  
 Oy 574 GCG 633  
 Db 541 GCG 600  
 Oy 634 ACCGAGAGATCATGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693  
 Db 601 ACCGAGAGATCATGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
 Oy 694 CACATGGCG 753  
 Db 661 CACATGGCG 720  
 Oy 754 CCGGAGCTGTGAGCG 813  
 Db 721 CCGGAGCTGTGAGCG 780  
 Oy 814 GCGCGCTGTGAGCG 873  
 Db 781 GCGCGCTGTGAGCG 840  
 Oy 874 CTGGAGCG 933  
 Db 841 CTGGAGCG 900  
 Oy 934 GCGCGAGCG 993  
 Db 901 GCGCGAGCG 960  
 Oy 994 GATGAGTGTGCG 1053  
 Db 961 GATGAGTGTGCG 1020  
 Oy 1054 GCG 1113  
 Db 1021 GCG 1080  
 Oy 1114 CCGGAGCG 1173  
 Db 1081 CCGGAGCG 1140  
 Oy 1174 GCGAGCTTGTGAGCG 1233  
 Db 1141 GCGAGCTTGTGAGCG 1200

Oy 1234 ACGGGCGATGGAAGATACCGGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1293  
 Db 1201 ACGGGCGATGGAAGATACCGGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260  
 Oy 1294 CAGCTGACG 1353  
 Db 1261 CAGCTGACG 1320  
 Oy 1354 ACCGCGACCTTGTGAGACCGCGCTGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413  
 Db 1321 ACCGCGACCTTGTGAGACCGCGCTGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
 Oy 1414 GCGCGCGCGCTGTGAGACCGCGCTGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1473  
 Db 1381 GCGCGCGCGCTGTGAGACCGCGCTGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440  
 Oy 1474 CGCAGCG 1488  
 Db 1441 CGCAGCG 1455

RESULT 6  
 AF332548  
 LOCUS AF332548 1455 bp DNA linear BCT 04-MAR-2001  
 DEFINITION Alcaligenes faecalis D-aminocyclase gene, complete cds.  
 ACCESSION AF332548  
 VERSION AF332548.1 GI:13194735  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Alcaligenes faecalis.  
 Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

REFERENCE  
 AUTHORS Hsu,C.S., Lai,W.L., Chang,W.W., Yang,Y.B. and Tsai,Y.C.  
 TITLE Gene cloning, nucleotide sequencing of Alcaligenes faecalis DA1  
 D-aminocyclase.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1455)  
 AUTHORS Hsu,C.S., Lai,W.L., Chang,W.W., Yang,Y.B. and Tsai,Y.C.  
 TITLE Direct Submision  
 JOURNAL Submitted (28-DEC-2000) Institute of Biochemistry, National  
 Yang-Ming University, 115, Li-Nong St. Sec. 2, Shih-Pai, Taipei,  
 Taiwan 11221, R.O. China

FEATURES  
 source  
 Location/Qualifiers  
 1..1455  
 /organism="Alcaligenes faecalis"  
 /db\_xref="taxon:511"  
 1..1455  
 /EC\_number="3.5.1.81"  
 /note="Daaf"  
 /codon\_start=1  
 /transl\_table=1  
 /product="D-aminocyclase"  
 /protein\_id="AAK15530.1"  
 /db\_xref="GI:13194735"  
 /translation="MSQPDATPEDYILSGTVIDGTNAPRLADVGVRGDRIRAAVGD  
 SASARRRIVAGRVVSPGFDISTHDNDLHKRDTPISGVTYVTCNGSILA  
 PLAAHPADLDLDEGSRFAFSYELAAAPPAVAAACMGHSTRRAVMDL  
 RREATADPEIOMOLADALASALGISTGCAFPRAAHSTETIEIVCRPLITGGVY  
 ATNHRDGEHIVALEETPRIGREDLVPIYISHKYGKLNFRSEKTELALAEAAVS  
 ODVSLDPRVPAVSATMLKODRVLLAGRTLLITWCKPYPELSDGDLIEIAERKSYDV  
 VPEIQPAQATYFMDEPDVRIILAFGTMTGSDGLPHDESPHRLMTGTPPVYAGHSR  
 DLGIFLETVMKRTGILTAKEFGLAERGQVPEYVADLVVFDPAIVADSAETEPHER  
 AAGIHSTYVGAAVMEQSPFGHARVYNRAA"

CDS

1..1455  
 /EC\_number="3.5.1.81"  
 /note="Daaf"  
 /codon\_start=1  
 /transl\_table=1  
 /product="D-aminocyclase"  
 /protein\_id="AAK15530.1"  
 /db\_xref="GI:13194735"  
 /translation="MSQPDATPEDYILSGTVIDGTNAPRLADVGVRGDRIRAAVGD  
 SASARRRIVAGRVVSPGFDISTHDNDLHKRDTPISGVTYVTCNGSILA  
 PLAAHPADLDLDEGSRFAFSYELAAAPPAVAAACMGHSTRRAVMDL  
 RREATADPEIOMOLADALASALGISTGCAFPRAAHSTETIEIVCRPLITGGVY  
 ATNHRDGEHIVALEETPRIGREDLVPIYISHKYGKLNFRSEKTELALAEAAVS  
 ODVSLDPRVPAVSATMLKODRVLLAGRTLLITWCKPYPELSDGDLIEIAERKSYDV  
 VPEIQPAQATYFMDEPDVRIILAFGTMTGSDGLPHDESPHRLMTGTPPVYAGHSR  
 DLGIFLETVMKRTGILTAKEFGLAERGQVPEYVADLVVFDPAIVADSAETEPHER  
 AAGIHSTYVGAAVMEQSPFGHARVYNRAA"

BASE COUNT 247 a 537 c 446 g 225 t  
 ORIGIN

Query Match 61.9%; Score 1088.6; DB 1; Length 1455;  
 Best Local Similarity 84.3%; Pred. No. 6.6e-110;  
 Matches 1226; Conservative 0; Mismatches 229; Indels 0; Gaps 0;  
 Oy 34 ATGTCCCAATCCAGTATCCAGCCCTTGCAGCTGTGTCCGCGGCGCGACCCATCATGAC 93

Db	1081	CCNACATGATCGGGGTCCGATGCCCTGCCCCAGACGAGCGCCCCACCCCTCGCCTGTGG	1140
QY	1174	GGCACCCTTCCCGCGGGGTCTGGGGACATATGCGCGGACCTGGGCTGTGTCCCGCTGGAG	1233
Db	1141	GGAACTGTTCCTCCCGGGTCTGGGGCACATCTCGCGGACCTGGGCTGTGTCCCATTTGGAG	1200
QY	1234	ACGGCGGATATGGAAGATGACCGGCGCTTGACCGCGCGCGCTTGCGCTGGCCGGGGCGGG	1293
Db	1201	ACCGCGGCTGTGGAAAAGACCGGCGCTGACCGCGCGCAATTGGGCTGGCCGAACGGGGC	1260
QY	1294	CAGCTGACGAGCGGGGTACTTCGCGCGACCTGGTGTGTTCGACCCCGGCACGGTGGCCGAT	1353
Db	1281	CAGGTGACGCCCGGCTACTACGCGCGACCTGGTGTCTTCGACCCGGGCACCGTGGCGGAC	1320
QY	1354	ACGCGCACCTTTCGACACCCCTACCGGAGCGCGCGCGGCGCATTCATTCCTGTACGTGAAC	1413
Db	1321	TGCGGACGCTTCGAGCATCCACCGAGCGCGCGCGGCGCATTCATTCGTGTACGTGAAC	1380
QY	1414	GGGCGCGCGGTCTCGCGAAGACGAGCGTTTCACCGGCGCGCATGCGCGCGGCGTGTGCGCA	1473
Db	1381	GGGCGCGCGGTCTCGGGAACACCACTGCTTCACTGGCCAGCATGCGCGCGCGCTGTCAAC	1440
QY	1474	CGCAGCGCGCGCTGA	1488
Db	1441	CGGCGCGGAGCCTGA	1455

```

/organism="Achromobacter xylosoxidans"
/isolate="A-6"
/sub_species="xylosoxydans"
/db_xref="taxon:85698"
l. 1497
/codon_start=1
/transl_table=11

```

/product-"N-acyl-D-amino acid amidohydrolase"  
 /protein\_id="BA08350.1"  
 /db\_xref="GI:1800111"  
 /translation="MDSRTLDAPADAFIAGATLIDGGGPARQGLAVGRIV  
 ALGDAHAGVIDARGLALAPFIDSTHDGILAPRELPTVSGITVYVNC  
 GIDSLPSRROIPDLPLPELPEFPAFEDMLRATAPAAVPIVGTTLARA  
 VMDYRATDAERARALDELLOAFGVTSTCTEYRPAAPATDELIDGPIRG  
 RAGIATYHLREADHYVAMEALLIGELDCRYFVSHKLAGENHRSRETLDM  
 SRAANORPCIDCHYRPASTYLRDLRALASRLITVSKYLPEDAPSEVMAELG  
 LDDEAIALRLAPAGIIFLMDADVNRIFSHPLTIVGSDLPDPHPRMGSTTNV  
 LRTMYREGLSLETAIHKMTGLAAOYGLTERGLIRGSHADLVLFDPANVTDTATF  
 SAPIVSGIHAHVNWNGVMDERTGERPQVLAIPGAPIMWSQSE"

BASE COUNT 235 a 592 c 467 g 203 t

Query Match 39.5%; Score 693.6; DB 1; Length 1497;  
 Best Local Similarity 67.7%; Pred. No. 6,3e-67;  
 Matches 988; Conservative 0; Mismatches 469; Indels 3; Gaps 1;

45 CGATTCCAGCCCTTCGACTGCTGCTGCGGGGACCCCTCATCGAGGACGACGAC 104  
 27 CGCTCCCGGCGAAGCGATTATATCCCGCGCCACCTCATCGAGGCGGGCGG 86  
 105 CCGGGGGGCGCGCGCGAGCTGGCGCGCGACCGCATCGCCATCGCGCATCT 164  
 87 CCGCGCGCGCGCGAGCTGGCGCGCGCGCGCATCGTGGCGCTGGCGCATCT 146  
 165 GTCGAGCGCGCGCGCGACACCGCGCGCGCATGTCGGCGCTGGTGGCGCGCTT 224  
 147 CGCCCATCGCTT 206  
 225 CATCGACTCGCGACACCGACGACGACGACGACGACGACGACGACGACGAC 284  
 207 CATCGACTCGCGACGACGACGACGACGACGACGACGACGACGACGACGAC 266  
 285 GATCTCGAGGCT 344  
 267 GGTTCGAGAGGACATACGAGGCTGTACCGCAATCGCGCATACGCGCGCT 326  
 345 GCGCGACGCT 404  
 327 GTCGCGGCGCGAGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 386  
 405 CGAGCGCTTCG 464  
 387 CGCCACCTTCG 446  
 465 CTGTATGTGGGCGCATTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 524  
 447 CGCGCTGTGCG 506  
 525 CACCGACGAGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 584  
 507 CACCGACG 566  
 585 CATGGGATTTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 644  
 567 CTTCGCGGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 626  
 645 CATCGAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 701  
 627 CATCGAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 686  
 702 CGAGGAGGCG 761  
 687 CGAGGAGGCG 746  
 762 GCGAGGCG 821  
 747 GCGAGGCG 806  
 822 GCGAGGCG 881

Db 807 CCGGAAAGCGTGGAGCATGATGACCGCGCGCGCGCGCGCGCGCGCGCGCG 866  
 882 GATTCCTTCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 941  
 Db 867 CCACCCCTCCCG 926  
 942 CACCATCATCAGCTGGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1001  
 Db 927 CACCGGATCAGCTGGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 986  
 1002 CCG 1061  
 Db 987 CATGGCGGAACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1046  
 1062 CATCTACTCATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1121  
 Db 1047 CATCTACTCATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1106  
 1122 GATCGGCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1181  
 Db 1107 GGTGGGCTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1166  
 1182 CCGCGGCTCGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1241  
 Db 1167 CACCAACCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1226  
 1242 ATGGAATGAGCG 1301  
 Db 1227 CCACAGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1286  
 1302 GCGCGGCTACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1361  
 Db 1287 GCAAGGCTATCATCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1346  
 1362 CTTCGAACCCCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1421  
 Db 1347 TTTTTCAGCTCCCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1406  
 1422 GGTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1481  
 Db 1407 GGTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1466  
 1482 CGCCTGAGCG 1501  
 Db 1467 TGCATATCTTGGAGTCAAC 1486

RESULT 8  
 AL646083  
 LOCUS 208050 bp DNA linear BCT 07-DEC-2001  
 DEFINITION Ralstonia solanacearum GM11000 megaplasmid, complete sequence;  
 accession AL646083  
 VERSION AL646083.1 GI:17431588  
 KEYWORDS Ralstonia solanacearum.  
 SOURCE Ralstonia solanacearum.  
 ORGANISM Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 Ralstonia.

REFERENCE 1 (bases 1 to 208050)  
 Salenow, M., Genin, S., Artiguenave, F., Guzy, J., Manganot, S.,  
 Arlat, M., Billault, A., Brothier, P., Camus, J.C., Catolico, L.,  
 Chandler, M., Choisme, N., Claudel-Renard, C., Cunne, S., Demange, N.,  
 Gaspin, C., Layle, M., Moisan, A., Robert, C., Saurin, W., Schlex, T.,  
 Siglier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M.,  
 Weissenbach, J. and Boucher, C.A.  
 Genome sequence of the plant pathogen Ralstonia solanacearum

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 208050)  
 Boucher, C.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
 Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie

COMMENT  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP277, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS,  
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,  
31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire  
INRA, BP27, 31326 Castanet-Tolosan Cedex  
Christian.Bouchet@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES  
source

1.208050  
Location/Qualifiers  
/organism="Ralstonia solanacearum"  
/strain="GM1100"  
/db\_xref="taxon:305"  
/plasmid="megaplasmid"  
135..947  
/gene="Rsp1115"  
/note="RS05505"  
135..947  
/gene="Rsp1115"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD18266.1"  
/db\_xref="GI:17431589"

CDS

gene

gene

CDS

/gene="Rsp1116"  
/note="RS05504"  
1162..2259  
/gene="Rsp1116"  
/EC\_number="4.1.2.5"  
/function="small molecule metabolism; degradation; amino  
acids"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PUTATIVE LOW SPECIFICITY L-THREONINE ALDOLASE  
PROTEIN"  
/protein\_id="CAD18267.1"  
/db\_xref="GI:17431590"

misc\_feature

gene

CDS

/note="ACUR or RS05503  
predicted by Framed"  
/evidence=not\_experimental  
complement(2462..3139)  
/gene="Rsp1117"  
/note="RS05502"  
complement(2462..3139)

gene  
CDSgene  
CDS

gene

CDS

/gene="Rsp1117"  
/function="miscellaneous; unknown"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PUTATIVE TRANSMEMBRANE PROTEIN"  
/protein\_id="CAD18268.1"  
/db\_xref="GI:17431591"  
/translation="MNDASLHETIISCFLEHOPRPTREIAERPNDEAGYRALRAL  
ADNHGYLHPHSDVWIAHPESAPTCYKSGRRKMGCMALGLVLAGGSAT  
ETRGADIDHVIIRIENGRLDDYVHFPIPKHADVITCSVOLFRDRAOYE  
WASRGIPKGDVRIEOWNFVAVETARHADADWTKMSVREAIELFARHLLSGPVWT  
AAEARE"  
complement(3188..3595)  
/gene="Rsp1118"  
/note="RS05501"  
complement(3188..3595)  
/gene="Rsp1118"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD18269.1"  
/db\_xref="GI:17431592"  
/translation="MDDKIAIVRBDLTWOKLNTAFIMSGITGASPEINGEAYRD  
ADNTHSICQVPVYLSDDQETLAVHRALEKVPGLYIEDMSTGHDAANRAVE  
AARGDSARVAVIALEREKIKYDKTGARMH"  
complement(3673..4161)  
/gene="Rsp1119"  
/note="RS05500"  
complement(3673..4161)  
/gene="Rsp1119"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD18270.1"  
/db\_xref="GI:17431593"  
/translation="MGNSGCEVCPARTIAVYGDRTVILIELFGYKREBEPOA  
OTGMSHSIRKLRLDAGIVVRLHLSDRPREYRLTSGDLPYLLSLKSGCK  
MGCFKATAPALVITTHRCGHETGLKIVCPACDEPFGPKATVTLGASFAERQARD  
GA"  
4228..4725  
/gene="Rsp1120"  
/note="RS05499"  
4228..4725  
/gene="Rsp1120"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"



Query Match 30.4%; Score 533.6; DB: 1; Length 208050;  
 Best Local Similarity 61.3%; Pred. No. 2.9e-50;  
 Matches 876; Conservative 0; Mismatches 549; Indels 3; Gaps 1;

gene  
 /protein\_id="CAD18271.1"  
 /db\_xref="GI:17431594"  
 /translation="MPATRRPREGAPRYLDFKVDLTETSGFAVTRDMILSPAEQY  
 DPQPMHLEADRYFGLVSGMOTLAIVMLRLDLRLGIPVAGAEIRDRFA  
 PMRGDYLRAAEVIAIRPSKSRPDRGMARVTTTNSGVTLYTQKMSLVPRRPF  
 QKVAE"  
 complement(5017..5610)  
 /gene="RS01121"  
 /note="RS05498"  
 complement(5017..5610)  
 /gene="RS01121"  
 /function="miscellaneous: not classified regulator"  
 /note="Product confidence: putative  
 Gene name confidence: hypothetical  
 predicted by Homology  
 predicted by Homology  
 /codon\_start=1  
 /evidence-not\_experimental  
 /transl\_table=11  
 /product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"  
 /protein\_id="CAD18272.1"  
 /db\_xref="GI:17431595"  
 /translation="MAEKITRRPRADAENRRLDVAATAPAEKVSASLIDIAE  
 AGVIGILYHFPTRDLIDETIRDEGRSLAEAROSEOPPLEAVRQMLLFVGL  
 ANKOIRADVINCMTDSCERLCTISGEVLITLQILIRAKOSGATGLAVEPDLISAV  
 AGVASFGEIDMNAEARLYEVAVAGLRVAATGGRK"  
 5822..6910  
 /gene="RS01122"  
 /note="RS05497"

gene

57 CTTCGACCTCTGCTGCTGCGGGGCGGACCTCTATCGACGAGCAACCCCGGGGCGG 116  
 95887 CATCCGCGAGGTCCTGCTGCTGCGAGCGGCGGCGGCGGCGGCGGCGGCGG 116  
 117 CGCGGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 176  
 95947 CGTCGCGATGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 96006  
 177 CGCGGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 236  
 96007 GGGCGAGGAGGAGTGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 96066  
 237 CACCGACGACGACACTACCTGCTCAGGCGTGCAGCATGACGCCCAAGATCTCGAGG 236  
 96067 CACCGACGACGACACTACCTGCTCAGGCGTGCAGCATGACGCCCAAGATCTCGAGG 96126  
 297 CGTCACCAAGGCTGACAGGCGCAATTGCGGCGATCAGCTGCGGCGGCGGCGGCGG 333  
 96127 CATCACCAAGGCTGACAGGCGCAATTGCGGCGATCAGCTGCGGCGGCGGCGGCGG 96186  
 354 CAACCG 413  
 96187 CGAGCG 96246  
 414 CGCGGACTACCTGAGCGCTTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 473  
 96247 CGCGGCTGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 96306  
 474 GGGCGATCAAGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 533  
 96307 CGGCGACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 96366  
 534 GGAATTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 593  
 96367 CGAAGTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 96426  
 594 TTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 653

Db 96427 GTCACCGGACGCGCTATGCGCAACGCCCATGCGCGCGCGCGCGCGCGCGCGCG 96486  
 QY 654 GTCCGCGCGCGCTGAGCGGCGCATGCGGCGCATCTACGCCACCCACATGCGCGAGGCGA 713  
 Db 96487 GCGCCAGCG 96546  
 QY 714 GCACATGCTGGCG 773  
 Db 96547 CGCCATCTGGAGGCGATGAGCGAGCGCTTACCGGCTGCGCGCGCGCGCGCG 96606  
 QY 774 GGTGATCTGCGACACCAAGATCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 833  
 Db 96607 CGTATCTGCGACCG 96666  
 QY 834 GCGGCTGATCGAGCG 893  
 Db 96667 GGAATGATCGAGCG 96726  
 QY 894 GCG 953  
 Db 96727 CGCGACGCTCG 96786  
 QY 954 CTGCTGCAAGCGCTTCCCGCAACTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1013  
 Db 96787 CTGCTGCGACCG 96846  
 QY 1014 CGCGCAATTCAGTACAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1073  
 Db 96847 GGGCGACGACCTGCTGCG 96906  
 QY 1074 GATGAGCAACCCG 1133  
 Db 96907 CATGCG 96966  
 QY 1134 CGGCTCTCG 1193  
 Db 96967 CGGCTCTCG 97026  
 QY 1194 GGGCGACTATGCG 1253  
 Db 97027 CGGCGACTATGCG 97086  
 QY 1254 CGGCTCTGACCG 1313  
 Db 97087 CGGCGACTGCG 97146  
 QY 1314 CGGCGACTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1373  
 Db 97147 GGGCGACTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 97206  
 QY 1374 TACCGAGCG 1433  
 Db 97207 CAGACGACCG 97266  
 QY 1434 CGAGCGGCTTACCG 1481  
 Db 97267 CGGTGCGCTGCG 97314

RESULT 9  
 ARA6  
 LOCUS  
 DEFINITION  
 accession  
 version  
 keywords  
 source  
 organism  
 reference  
 authors

1467 bp DNA linear BCT 10-FEB-1999  
 Alcaligenes xylosoxydans DNA for N-acetyl-D-aminic acids  
 amidohydrolase, complete cds.  
 D50061.1 GI:769683  
 N-acetyl-D-aminic acids amidohydrolase.  
 Alcaligenes xylosoxydans (strain A-6) DNA, clone pABD.  
 Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 Achromobacter.  
 1 (bases 1 to 1467)  
 Wakayama, M.

TITLE Studies on structure and function of bacterial N-acetyl-D-amino acids  
amidohydrolase  
JOURNAL Theses (1995) \_FACULTY\_/Applied Chemistry/\_MAJOR\_, Oita University  
REFERENCE 2 (bases 1 to 1467)  
AUTHORS Wakayama, M.  
TITLE Direct Submission  
JOURNAL Submitted (06-MAR-1995) Mamoru Wakayama, Oita University, Applied  
Chemistry, Danomaru, Oita, Oita 870-11, Japan  
(Tel:0975-69-3311(ex.746), Fax:0975-69-7957)  
FEATURES Location/Qualifiers  
source 1. 1467  
/organism="Achromobacter xylosoxidans"  
/strain="A-6"  
/db\_xref="taxon:85698"  
1. 1467  
/codon\_start=1  
/transl\_table=1  
/product="N-acetyl-D-amino acids amidohydrolase"  
/protein\_id="BA08778.1"  
/db\_xref="GI:1707291"  
/translation="MOEKLDLVIEGWVIDLGGPRRRADYVIGERIALTSDLSAAP  
ADRLDAGRIYAFIDTGHDDLMFVEKELMKTSOGITSVYVCGISGAPL  
PGNTAAALALIGDPLFADAMATRGALAEADPMNVALVGHANRLAAMPDPAOS  
AKBORAMERMDALDLAGAVSTGLAYOPGVAEQELDGLAVAAKGLHLSHR  
NEGVAVAEDEVTAAGRTGCAIVLSHKCMFAMNGKSAATLANIDRAAGVDVA  
LDITPFGSSSTILPERADJIDIRITWSPHPBCGQSLAEIARNGCDVTAARL  
CPGAIYFADENENVRIFQHECCMGVSDIPNAPHVPDADYVDMGTGTYLGRYVEAEI  
LLEAAVAKMTALPARVFGALDGRGLAVGAMADYVVDADTVDCAITWDAFTLASGI  
EHVYVNCAYFPQAPSPHRRGRIRRDASTAGAEFSR"  
BASE COUNT 213 a 530 c 530 g 194 t  
ORIGIN

Query Match 25.9%; Score 455.8; DB 1; Length 1467;  
Best Local Similarity 59.3%; Pred. No. 4.8e-41;  
Matches 813; Conservative 0; Mismatches 552; Indels 6; Gaps 2;

59 TCGACCTGCTCTCGCGGCGCGCACCCCTCATCGACGACGACACCCCGGCGCGCG 118  
14 TGGATCTGGTATGAAAGCGCGCTGGGTGATCGACGGCTGGCGCGCGCGCGCGCG 73  
119 CGACCTGGCGCGCGCGCACCATCGCGCATCGCGCATCTGTGCGACCGCGCG 178  
74 CGACGTGGCATCGGTGGCGACGCGCATCGCGCATCGGTGGCGCGCGCGCGCG 133  
179 CGACACCGCGCGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 238  
134 CGGACCG 193  
239 CGCAGACGACCACTACTGCTCAGCGGTGCGGACATGACGCCCAAGATCTCGAGGCG 298  
194 GCCACGATGACCTGATGTTGTGTAACAAACCGCGCGCTGGAATGGAACGACGAGGCA 253  
299 TCACACGCTGTGTCAGGGCAATGGGCAATGAGCGCGCGCGCGCGCGCGCGCGCA 358  
254 TCAGTGGTGTGTGGGCACTGCGGCACTGCGGCGCGCGCGCGCGCGCGCGCGCA 313  
359 CG 418  
314 ATACCG 370  
419 ACTACCTGGACCGGTGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 478  
371 TGTACTTGGCG 430  
479 ATTCAACGCTGGCG 538  
431 ACGCCAACCTGGCG 490  
539 TCGCGGCATGCGCGGACCTGCGCGGACGACGACGACGACGACGACGACGACGACG 598  
491 AGCGGGGATGGAGCGGATGCTGGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 550  
599 CGGCGCGCTTCTACCG 658

551 CGGGCTGGCTTACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 610  
659 GCGCGTGAAGCGCGCATGGCGCGCATGTACGCCACCAATGCGCGGAGCGGACGACA 718  
611 GCGTGGCG 670  
719 TGTGGCGCGCGCGTGGAGGAACCTTCCGATGGCGCGCGCGCGCGCGCGCGCGCGCG 778  
671 TCGAGCG 730  
779 TCTCGCACCAAGGATGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 838  
731 TGTCCGACCAAAATGATGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 790  
839 TGTATGA--GCGCGCATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895  
791 ACATGACCG 850  
896 CCGGCTCCACCATGCTCAAGAGACCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 955  
851 GAGTTCCACCATCTGTATACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 910  
956 GGTGCAAGCCCTTCCCGAATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015  
911 GGTGCAAGCCCGATCCCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970  
1016 GCAATTCAGATGACGCTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075  
971 GCTGCGACGCGGTACG 1030  
1076 TGGACGACCG 1135  
1031 TGGACGACGAGGTGCG 1090  
1136 GCTGCG 1195  
1091 GCTGCG 1150  
1196 GGCATATGCG 1255  
1151 GCGCTAGTGTGCG 1210  
1256 GCTGACG 1315  
1211 GCGTGGCG 1270  
1316 CCGACCTGTGTGTTGACCG 1375  
1271 CCGACGTGTGTGTGACCG 1330  
1376 CCGAGCG 1426  
1331 CGTGGCG 1381

RESULT 10  
LMEFLCHR34\_10/c  
WPCOMMENT  
Sequence split into 18 fragments LOCUS LMEFLCHR34 Accession AL499623  
Fragment Name Begin End  
LMEFLCHR34\_00 1 110000  
LMEFLCHR34\_01 100001 210000  
LMEFLCHR34\_02 200001 310000  
LMEFLCHR34\_03 300001 410000  
LMEFLCHR34\_04 400001 510000  
LMEFLCHR34\_05 500001 610000  
LMEFLCHR34\_06 600001 710000  
LMEFLCHR34\_07 700001 810000  
LMEFLCHR34\_08 800001 910000  
LMEFLCHR34\_09 900001 1010000  
LMEFLCHR34\_10 1000001 1110000  
LMEFLCHR34\_11 1100001 1210000



small overlap between neighbouring submissions. Cosmid 2K36 overlaps cosmid SCK15 and cosmid SCK13.

FEATURES  
source  
Location/Qualifiers  
1..22251  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid 2K36"  
3..112  
/note="nominal overlap with streptomyces coelicolor cosmid SCK13"  
complement(160..864)  
/gene="2SCK36.01c"  
/note="SCO4978"  
complement(160..864)  
/gene="2SCK36.01c"  
/note="2SCK36.01c, conserved hypothetical membrane protein, len: 234 aa; similar to SW:YAB5\_MCTU 26.0 kDa protein RVI085c or MYV017.38c, 242 aa; fasta scores: opt: 563 Z-score: 653.6 bits: 128.3 E(): 8.6e-29; hydrophobic membrane spanning regions"  
/codon\_start=1  
/transl\_table=11  
/product="putative integral membrane protein"  
/protein\_id="CAC40591.1"  
/db\_xref="GI:14285272"  
/db\_xref="SPTREMBL:093JL6"  
/translation="MTASVPDAPRDLPAGRGPTVLSLPHYKKRLKGLHLMGMPAYLVAGLVLAETSEARVACGVYVLTACILGVSALYHRGDMSPMDVLRDLANI FLIAGTPTLMLLPGAKSEHMLMGWMAAAGIVFRVWVGAPRWLYPCYILAMG MAAYVLEPRRAGGVAVLVYVGLLYSAGVYLYGKRRNPSPRMGFREYHSPT 1141..1145  
1151..2980  
/gene="2SCK36.02"  
/note="SCO4979"  
1151..2980  
/gene="2SCK36.02"  
/note="2SCK36.02, probable phosphoenolpyruvate carboxykinase, len: 609 aa; similar to SW:PECK\_CHLII (EMBL:556812) Chlorobium limicola phosphoenolpyruvate carboxykinase [GFP] (EC 4.1.1.32) PCKA, 646 aa; fasta scores: opt: 2400 Z-score: 2689.0 bits: 507.7 E(): 3.6e-142, 58.3618 identity in 598 aa overlap. Contains pfam match to entry PF00821 PEPCK, Phosphoenolpyruvate phosphoenolpyruvate carboxykinase (GTP) signature"  
/codon\_start=1  
/transl\_table=11  
/product="putative phosphoenolpyruvate carboxykinase"  
/protein\_id="CAC40592.1"  
/db\_xref="GI:14285273"  
/db\_xref="SPTREMBL:093JL5"  
/translation="MARDAIAPVPVNNHOLISWNEIALTOPDAVWVWDCSEAEYE RLCELYEKGTPRKIDPYKRNSTYAAADPTVAVEDTFLCSAKEEAPGNTNMDK RMTAKGIVDELDEDEGEFVAVASVGALEPGADVPNCSNRTKYSHPFEDNEM STGSGYGGNALLGKRCYALRIASVVARDEGMALAHMLYKLPPTGARGVAAAPSA CKRTNLAILEPTISGMTVEITGDDIANKMRGDSGLIYINPAGFFVAPGEGHTNA FTTPAACQPIIAPEWDEPRGPVISAILEGRRATVAPLVTSFDNNGVFLGANVASE KTAAGKVGKELRRPFFMLPCGNGMDYGHWDVAKDKDSKLPKITYVNNFRKD DAGRTWPGFGENGVLMVIERLEGRADYETPIGVLPTEKESLDTDLADLADLER ILSDVKEVWREELVPEHLNTEFGDHPALMDYRALVRLG"  
1205..2974  
/gene="2SCK36.02"  
/note="Pfam match to entry PF00821 PEPCK, Phosphoenolpyruvate carboxykinase, score 1344.70, E-value 1967..1993  
/gene="2SCK36.02"

gene  
/note="PS00505 phosphoenolpyruvate carboxykinase (GTP) signature"  
complement(3263..4213)  
/gene="2SCK36.03c"  
/note="SCO4980"  
complement(3263..4213)  
/gene="2SCK36.03c"  
/note="2SCK36.03c, unknown, len: 316 aa"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein 2SCK36.03c"  
/protein\_id="CAC40593.1"  
/db\_xref="GI:14285274"  
/db\_xref="SPTREMBL:093JL4"  
/translation="MSGDLEGLAIVNRITLRQMRGPPYCGAGTGHVCLVPAGA GLPGLIEAIGSRGEPRLRVVTGTTTPAAGSAGLPLRTPEGRTVYMRWAGDRKIG CGTVRADDDAGVAPVYLVAEADPAAGASAPATWDGIVATGRTAARARVMDPVS EALRGLPLGDKYRLVIEFGAGAFGYOLQVPRDHDSTDLVRTTELMGEMARTGV SLMEPPYVAPAGGLQWASRWDGDFYRLVGGDPDPVPLANEVDPSWERFDGP TGEFVRMLTEPDHPSTARNVDVHTYDYGPAADDEGV"  
complement(4210..5610)  
/gene="2SCK36.04c"  
/note="SCO4981"  
complement(4210..5610)  
/gene="2SCK36.04c"  
/note="2SCK36.04c, possible bifunctional protein, len: 466 aa; N-terminal region presents low similarity to N-terminal region of TR:O14362 (EMBL:297992) Schizosaccharomyces pombe putative glucan synthase-1 (EC 2.4.1.34) SPBC30D10.17c, 504 aa; fasta scores: opt: 176 Z-score: 188.3 bits: 44.3 E(): 0.0071; 24.931% identity in 361 aa overlap and C-terminal region to C-terminal region of TR:AK22951 (EMBL:AE005774) Caulobacter crescentus oxidoreductase, GFO/IDH/MOA family CC0967, 307 aa; fasta scores: opt: 146 Z-score: 160.2 bits: 38.4 E(): 0.26; 26.590% identity in 173 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="putative bifunctional protein"  
/protein\_id="CAC40594.1"  
/db\_xref="GI:14285275"  
/db\_xref="SPTREMBL:093JL3"  
/translation="MSIHDFATWELRLRYRASHPERLAGPCHVWGQIGAGWSVP GRVPHAPGRASVEDMODEFTVEAQAALRAHQOSVSEFVETVMDGTLAAVVDV GRVPHAPGRASVEDMODEFTVEAQAALRAHQOSVSEFVETVMDGTLAAVVDV ANGATRAETRAEOTRLGVLTPDELKALRYRVARMDMDGDAVYAOGRVDAVCCCLP LQGLYADARSRCMLQGFATDAVAPDAVAVOGVLSGPGWIAFGDGGDRLADLT DPVSAAVARELEVILVVGREGEPRLAPAGLPRLTLHACGALADPLEIGLTRLE YLRGPEERVELLDGAVPRSLAAHIEVGEHHPRIIALLNELSLMDRPLISRTV LEGDIDTDTAGGAR"  
complement(5618..5621)  
5786..5792  
5801..5983  
/gene="2SCK36.05"  
/note="SCO4982"  
5801..5983  
/gene="2SCK36.05"  
/note="2SCK36.05, putative membrane protein, len: 60 aa. Contains possible hydrophobic membrane spanning region"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="CAC40595.1"

Query Match  
Best Local Similarity 9.8%; Score 171.8; DB 1; Length 22251;  
Matches 813; Conservative 0; Mismatches 647; Indels 183; Gaps 7;  
OY 32 CCATGCCCATCCGATCCAGCCCTTCGACTGCTGCGCGGCGGACCCCTCATCG 91  
DB 10533 CCGGAGCCAGAGAGAGTCCAGCATGGAAGAGCTGTCATCAGGACCGCAGCTCTG 10474

misc\_feature

misc\_feature

QY 92 ACGGAGCAACACCCGGGGGGGGGGGACCTGGGCTGCGCGGCGACCCGATCGCC 151  
 DB 10473 ACGGCTCGGGGGGGGACCTCTACCGAGCCGAGCTGTGGTCGACGGCGGCGGATGCT 10414  
 QY 152 CCATCGGAGATCTGTGGAGACGGCGCGGCGACACCCG-----GTCG 193  
 DB 10413 CGATGCTCAAGAGAGCGGGCGGGCGGGGTGGCAGCGGGCGGACGGCGGAGCTGG 10354  
 QY 194 ACGTGTGGGCTGTGGTGTGGCGCGCGGCTTCATGACTGCGACACCGACGACAACT 253  
 DB 10353 ACGCGAGGGGCTGTGCTGTGTCGCCGCTTCATGACATGACACCGACGACGCTGG 10294  
 QY 254 ACGTGTGAGCGTGGGCGACATGAGCGCCAAAGATCTGCGAGGGCTTCACACAGGTGTCA 313  
 DB 10293 CGGTGCTCGGGAGACCGGACCAACAGCGCCAAAGCGCCCGGCGGCTTCACACCTGGAGGTCA 10234  
 QY 314 CGGGCAATTCGGGCAATGACCTGGCGCGCTGGCGGCGACCGCAAC----- 358  
 DB 10233 TCGGGCGAGAGCGGGCTGTCTACGACCGGTGACGACCGGACCGCTCGGCGAGTGGCG 10174  
 QY 359 ---CGCGCGCGCGCGCTGACCTGTGGAGCAAGCGGCTCTTACCGTTTCAGCGCTTCG 415  
 DB 10173 GGGCGATGCGGGGTGGAGACGGCCACGGGAGAGAGTGTGACTTCGACTGGGCTGCTG 10114  
 QY 416 CCGACTACTGAGCGCGTGGCGGCGACGCGGCGCGCTCAAGCGCGCTGTATGG 475  
 DB 10113 GCGAGTACTGTGAC---CGGCTGACCGGGGCGATCGCGGTCAAGCGCGCTGATGCTC 10057  
 QY 476 GCCATTACGCTGCGCGCGCGCGCTCATGCGGACTTGGACGCGCGCGCCACCGGAGG 535  
 DB 10056 CGCGGAGCGGTGCGCGCGCTTCGCGCTGCGGAGGACCGGAGGCGGAGCGAGGG 9997  
 QY 536 AATTCGCGGCGATCGGGAGACCTGGCGGAGAGGACCAATGAGCGGCGGCTTCGCAATT 595  
 DB 9996 AGCTGAGACATCTGCGCGCGCTGTGCTCGAGGCGCTGGAGCGGCGGCTGCTGT 9937  
 QY 596 CGACCGCGCGCTTCTACCGCGCGCGCGCGCGGCGGCGACCGAGAGATCATCGAGGT 655  
 DB 9936 CGTGGGCGGCTGACTTACGCGCGCGGAGTACGCGCAAGAGCGCGGAGCTGACGAGCT 9877  
 QY 656 GCGCGCGCGGAGCGGCGGCGGCTCATGCGCCACCGACATCGCGGAGCGAGCGGAGC 715  
 DB 9876 GCGGCGGTGGGTGCTGTACGCGGCGCTACTGCGCGGCGGCGGCGGCGGCGG 9817  
 QY 716 ACATCGTGGCGGCTGGAGGAAACCTTCGCGCATCGGCGCGGCGGCTGAGCTGCGG 775  
 DB 9816 GGGCGCTGAGGCGGTACGAGAGAGTGTGAGCTGACCGCGGAGCGGCGGCTGCGG 9757  
 QY 776 TGATCTGCGACACAAAGTCTATGGCGGCGGCGGCGGCTGCGGAGAGCGTGC 835  
 DB 9756 ACCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9697  
 QY 836 CGGTGATCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 892  
 DB 9696 CCGTGTGCGAGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9637  
 QY 893 TGGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 913  
 DB 9636 CCGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9577  
 QY 914 AGCAGACCGCGGCTGTGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 973  
 DB 9576 AGAGATCTCTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9517  
 QY 974 AACTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1003  
 DB 9516 AGGTATGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9457  
 QY 1004 ---CGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1033  
 DB 9456 CCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9397  
 QY 1034 ---TGCTGCCCGGAGCTGCGAG 1051

DB 9396 GCGTGGCGGCGGAGAGCCCGGTGAGAGTGTGCGCGGCGGCGGCGGCGGCGG 9337  
 QY 1052 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1111  
 DB 9336 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9277  
 QY 1112 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1171  
 DB 9276 GCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9217  
 QY 1172 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1231  
 DB 9216 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 9157  
 QY 1232 AGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1291  
 DB 9156 AGAGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9097  
 QY 1292 GCGAGCTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1351  
 DB 9096 GGTGTGTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9037  
 QY 1352 ATACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1411  
 DB 9036 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 8977  
 QY 1412 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1471  
 DB 8976 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8917  
 QY 1472 CACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1494  
 DB 8916 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8894

RESULT 12  
 AF453501/c  
 LOCUS  
 DEFINITION  
 agent ansamitocin biosynthetic gene cluster I, partial sequence.  
 ACCESSION  
 AF453501.1  
 VERSION  
 AF453501.1 GI:21449342  
 KEYWORDS  
 Actinosynnema pretiosum subsp. auranticum.  
 Actinosynnema pretiosum subsp. auranticum.  
 Actinosynnema pretiosum subsp. auranticum.  
 Actinomyces; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Pseudonocardiales; Actinosynnemataceae;  
 Actinosynnema.  
 1 (bases 1 to 82746)  
 Yu, T.-W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K.O.,  
 Xu, J., Moss, S.J., Leister, E. and Floss, H.G.  
 The biosynthetic gene cluster of the maytansinoid antitumor agent  
 ansamitocin from *Actinosynnema pretiosum*  
 Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7968-7973 (2002)  
 JOURNAL  
 MEDLINE  
 22056096  
 PUBMED  
 12060743  
 REFERENCE  
 2 (bases 1 to 82746)  
 Yu, T.-W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K.O.,  
 Xu, J., Moss, S.J., Leister, E. and Floss, H.G.  
 Direct Submission  
 Submitted (28-Nov-2001) Department of Chemistry, University of  
 Washington, Box 351700, Seattle, WA 98195-1700, USA  
 FEATURES  
 source  
 1..82746  
 /organism="Actinosynnema pretiosum subsp. auranticum"  
 /strain="ATCC 31365"  
 /sub\_species="auranticum"  
 /db\_xref="ATCC:31365"  
 /db\_xref="taxon:42198"  
 <1..82653  
 /note="cluster I of the maytansinoid antitumor agent  
 ansamitocin biosynthetic gene cluster"

gene  
complement (4854. .5057)  
/gene="asm32"  
/complement (4854. .5057)  
/gene="asm32"  
/codon\_start-1  
/transl\_table-11  
/product="unknown"  
/protein\_id="AA054110.1"  
/db\_xref="GI:21449378"  
/translation="MITEQNARIIPARRSSGSRNARCAVWMPRPQPSRATRAPGPRS  
SAANDSRARSGMDPRTSLNRSA"  
5415. .6461  
/gene="asm31"  
5415. .6461  
/gene="asm31"  
/codon\_start-1  
/transl\_table-11  
/product="RNA polymerase sigma factor"  
/protein\_id="AA054109.1"  
/db\_xref="GI:21449377"  
/translation="MWTGSADGRRRRPVKKMGVTEAFWETDEDIKLAATLRDNEA  
ANABDSTRVLAIGRVRLTLADQEVELARVEAGLVAEHLDLPPAARSRRPVLR  
RVVRGQRKHLDLANRLVYSIAKRTKQGMALDLVIOEGNLGVRAEKEDYV  
FESVYATWIMRQIATRLADQSRTIRLPVHMYEVTKLARIENDLHNRGRATAEE  
VVEVDLYTEKEVVELRGHARTPSIDORTIGAEVDTSIGDILADENNAVPLEVSSLL  
SMDLRVLETTLREAGVVRNRRGCPGCVLHTLDEVGRTVGTTRKRLIESTRTAKL  
OHPARSGICEYLD"  
complement (6501. .9518)  
/gene="asm30"  
complement (6501. .9518)  
/gene="asm30"  
/codon\_start-1  
/transl\_table-11  
/product="cytochrome P450"  
/protein\_id="AA054108.1"  
/db\_xref="GI:21449376"  
/translation="MVAATGRIRIPGPKPLPLVGNLLDVLTSDDLTQDVDFLDRCHREGG  
LYALTFACOROVFSHELVARMCSDPSMCKAAVHAPALRGVDRAGGLTEAGDSDPW  
GAHNLPLAPGPTAMRDHPFAMLDIAEOMLVARRRGDPDRITDADMDKTRLLDTPIA  
LCAFAFARNSSYRDRAPHFVDVAWVRSYVEAGSEALRQGPPTLVGNRRORYDDIATM  
NRINAGIYAAARALLPAGRPDDLLERLTCADPYTGERLSARVRITQLVETLIAGHET  
TSGILSFVHRLAHPEYLKRAKDANDVGDVPAFEDRLDYLGQVLRTELRLHP  
TRAPALLAPDEPAELGGHAGAGEPVLVMLPTLHRPAAWRDPDFDEPRFAPEBDE  
IPACAMWFFGGARACIGRPEALQEAFLVALVQLQFDLADDPDHRLITKOTILKP  
DSLVRAPRPRDGCATATETVPHQVPATHRGTPLHVFYSGNGSGEGALRTIAG  
DGAAGVATSVAPLDDAVRALPASPGPVIVSSYNGAPDNAAHFYRRTITQGPPLSG  
VDYLVLGGNGLDMSATQVRYPTLIDEMAAAGRRIRREKATDADADFGDEMYTER  
LMPILSSACGVEGEISPTRRVYESDADAGLGSAAVLENRELVKRPDAGSKHLE  
LRLDGTSYRIGDYLSTLPONHPDLVRLVAVRIGTAEVYVESASPTGVPVGRAL  
RVDELTLKCDLSPAGAVVARIABRCQCPERARLALATGATLLELLEPSCAVD  
LALNELPVRRLYSISSAAEORAEVALVYSVTGVSLSHRPGDVAVAGIAS  
PPESEFRPADNTVAVVLIAGTGIAPFRGFLRABARLAGEEPALLIFGCGPELIDLI  
YAEFAPLADMLLEVDRAYSRHPDGEVRHVOHRLRDRRVRVLYDAGARYILCGDATR  
VGPVPEVLEIRIGHGAGWIDALAGRGATYATVF"  
9664. 10245  
/gene="asm29"  
9664. 10245  
/gene="asm29"  
/codon\_start-1  
/transl\_table-11  
/product="transcriptional regulator"  
/protein\_id="AA054107.1"  
/db\_xref="GI:21449375"  
/translation="MSAAVRREOLVAAALRVAKRKGIDIAAATTTRICAEADPBGAFHY  
CFRKSGLYALLTGLINTSIDAMPALVSPDLPAVDNRTMLAEPWAGDPAELV  
LGEVLTLAQEPSTRELPAHNRREYAAVNAHLDLAEAGFDIYDRLALAEVLVSV  
LDGVAWSWFHRBDELAKSLDQFAVRRALTR"  
10393. 10857  
/gene="asm28"  
10393. 10857  
/gene="asm28"  
/codon\_start-1



```

/transl_table=11
/product="unknown"
/protein_id="AA54106.1"
/db_xref="GI:21449374"
/transl_table="MTDTTNRHTPVONALAVGAVFLVINGEYVPEGLTNTYDLEF
AGHSDAALLGKRVKVALNHLVHLAGAGIAGLSAPGAGAFLLGGAIYAVLWL
YGVYVDEHSTANVPLNADNMLHLGLAVGMIGLGAALGRNHAHYRAASDD"
/complement(10912..11937)
/gene="asm27"
/complement(10912..11937)
/codon_start=1
/transl_table=11
/product="glucose 1-dehydrogenase"
/protein_id="AA54105.1"
/db_xref="GI:21449373"
/transl_table="MPRLAEQVYVVVSSGIGRYTALFPAAGARVVCANALRAL
DLVEVRGAGRAVAVTADIDAEAAVRADVAVRFGVDVWVNAAGVGVYVD
TFAGEFDRVRYVYLGHVHAKALPALRRAGGVLIQVAVSLGASAPLQAPYAAKS

Query Match      6.6%; Score 116; DB 1; Length 82746;
Best Local Similarity 44.2%; Pred. No. 0.00011;
Matches 758; Conservative 0; Mismatches 935; Indels 21; Gaps 6;

QY 35 CATTGACCTGCTGCTGCGCGGCGGACCTCATGACGCGAGCAACCGCGGGGGGC 115
Db 73869 CCGACACCTGCTGCTGCGCGGCGGAGCTGCGCGCGCGCGCTGCGAGCTGCTGCG 73810
QY 116 GCGCCGACCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 175
Db 73809 CCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73750
QY 176 CCGCGGACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 235
Db 73749 CCGCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73690
QY 236 ACACCCGACGACGACAACTTACTGCTGAGCGGTGCGGACGATGAGCGCGGCGGCGG 295
Db 73689 ACCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73630
QY 296 GCGTACACGAGGTGCTGACGAGGCGAATTGCGGCGATGACGCTGCGGCGGCGGCGG 355
Db 73629 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73576
QY 356 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 415
Db 73575 TCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73516
QY 416 CCGACTACCTGAGCGGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 475
Db 73515 CCGACCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73456
QY 476 GCCATTGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 535
Db 73455 GCCAGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73396
QY 536 AATATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 595
Db 73395 CCGCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73336
QY 596 CGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 655
Db 73335 CCGCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73276
QY 656 GCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 725
Db 73275 CCGCGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73216
QY 716 ACATGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 772
Db 73215 TCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73156
QY 773 TGGTATCTCGGACGACGACGATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 832

```

```

Db 73155 CCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73096
QY 833 TCGCGCTGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 892
Db 73095 AACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73036
QY 893 TCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 952
Db 73035 GCGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72976
QY 953 CCTGTGTGAACCTTCCCGCACT---GACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1009
Db 72975 CTTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72916
QY 1010 AGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1069
Db 72915 GCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72856
QY 1070 TCATGATGACGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1129
Db 72855 TCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72796
QY 1130 CCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1183
Db 72795 GCGAGCTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72736
QY 1184 CCGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1243
Db 72735 TCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72676
QY 1244 GGAATGATACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1303
Db 72675 TCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72616
QY 1304 CCGGCTGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1363
Db 72615 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72556
QY 1364 TCGAACACCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1423
Db 72555 TGGGTTC---GACACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72499
QY 1424 TCTGCGGACGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1483
Db 72498 CCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72439
QY 1484 CTTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1543
Db 72438 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72379
QY 1544 CAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1603
Db 72378 TCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72319
QY 1604 ATATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1663
Db 72318 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72259
QY 1664 TGAACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1723
Db 72258 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72199
QY 1724 AGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1757
Db 72198 CCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72155

RESULT 13
SC6A11 28751 bp DNA linear BCT 12-MAY-2002
LOCUS Streptomyces coelicolor cosmid 6A11.
DEFINITION AL138978 AL645882
ACCESSION

```

Nominal overlap with Streptomyces coelicolor cosmid 8F4 complement(6. .872)

2103. .2107  
2118. .3143

```

/note="SC05426; pfkA2"
2118. .3143

```

```

/gene="SC6A11.02" pfkA2, 6-phosphofructokinase 1 100 244
/note="SC6A11.02,

```

6-phosphofructokinase from *Amycolatopsis methanolica* (35 aa) fasta scores; opt: 1274, z-score: 1463.6, E(): 0

59.1% identity in 342 aa overlap. Also similar to  
 TRNEM: CAB51967 (EMBL: AL109661) Pfk, 6-phosphofructokinase

from *Streptomyces coelicolor* (342 aa) fasta scores; opt: 1251, z-score: 1437.4, E(): 0, 56.78 identity in 342 aa

Phosphoinositide 3-kinase and phosphoinositide-dependent kinase-1 (PDK1) signature" .

```

/db_xref="GI:6941668"
/db_xref="SWISS-PROT:Q9J1L8"

```

DEAESVF# .  
2121. .2804

Phosphofructokinase, score 278.20, E-value 1.6e-101\*

phosphofructokinase, score 52.20, E-value 9.9e-19  
913. .2969

note="PS00433 Phosphofructokinase signature"  
complement(3153..3866)

```
note="SC05427"
complement(3153, .3866)
```

complement(3978, 4868)

```

note="SC05428"
complement(3978, 4868)

```



gene  
CDS

gene  
CDS

gene  
CDS

```
/note="identified by Glimmer3; putative"  
/codon_start=1  
/transl_table=11  
/product="hypothetical protein"  
/protein_id="AAK23406.1"  
/db_xref="GI:13425162"  
/translation="MYLISAHKCEGSPVSGFNAVLSCGMYLAAPAPCARODRTRRPDD  
WGGRHRRSRGSTRNRSGPAPAPAOAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP  
TVAPVPRGRTTDPFTVRLDVLDTVCCKPQVAGDFALVSYGKKRKDEQWALERK  
PNNFTLDNGSNKNKCTVYITIDYSCAPEDAOELANLHDMAWMSPOLRLLRNDQTVG  
SPFRRTVSDMAANAGGAGVGLTMRKLRLDFTSVGKNFERAOIILYSTTSR"  
complement(1018..1278)  
/gene="CC3445"  
complement(1018..1278)  
/gene="CC3445"  
/note="identified by match to protein family HMW"  
/codon_start=1  
/transl_table=11  
/product="ATP synthase F1, epsilon subunit"  
/protein_id="AAK23407.1"  
/db_xref="GI:13425163"  
/translation="MAKLHSLVAPEREELFSGEVDVYQACAGECGVLANHPAEMTWT  
LRGEVAVYKDAATKRVFDIGGFAVDVPEGLITLAENAVZAA"  
complement(1278..1529)  
/gene="CC3446"  
complement(1278..1529)  
/gene="CC3446"
```

CDS

**gene**

**CDS**

gene

```

/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK25406.1"
/db_xref="GI:13425162"
/translation="MWLISHKSCGEGSVGFNAVLVSSCGYLAAPCAKRODRTRPND
WGRNREFRRSGNSGNRSGSPAPAQAQAPAPAAQAAPADAAVAAPAEATPAD
TVAPAVFPTTGPFPVRLDYLDPAARAPAKAPAGVGADLVSGFKRKREQWALKE
PENTVDLPDSNNKTCTVTIDTSQAEQDLANGLHMAAWENSPQRLRIINDQTGV
SDRRFTVSHDDAMAGRGVLVRLMKLIDETSVKNFEERAQILYSTSR"
complement(1018..1278)
/gene="CC3445"
complement(1018..1278)
/gene="CC3445"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, epsilon subunit"
/protein_id="AAK25407.1"
/db_xref="GI:13425163"
/translation="NAKHIFGSIVAPERLEFGEVDYTAQAPCAESDFGVLANHPFTT
LRGGKYTVGDGATTFVPFDGSEFADVGPBGGLTLIAEHAENA"
complement(1278..1529)
/gene="CC3446"
complement(1278..1529)
/gene="CC3446"
/note="identified by GLIMMER2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK25408.1"
/db_xref="GI:13425164"
/translation="MTEDADIANCCEDELAENVSPANGLIHDHGILLMLGEREE
EGEVGYATIRFCPEGTGETOEDIEMALGVETFN"
complement(1529..3148)
/gene="CC3447"
complement(1529..3148)
/gene="CC3447"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, beta subunit"
/protein_id="AAK25409.1"
/db_xref="GI:13425165"
/translation="MKTPEKPRATAKKRAAPKAAPAAAAPKAAPAAKAPKP
AFAPAAPAPAPAGLEGATGRVVOITIGAVDIIEFGALPALIANAVENTATQRVY
EFAQHIOGSTVRALAMDTEGIVNRQPRNDGIRVYPCPGTIGRMANTGEPIDE
TYMOELINNIACVAGCVSLANGERRREGNDLHEMIENPVNDPKANNSTGESR
CAYVQENMPBARARARVALTGISTAEFRDEGVDLLFYDNIFRFQAGAESALL
ATTVSARDIAADIPVADPDSTSRINDPLVIGEHTVARRVQEVLYDKTKADII
AIIQMDLESBEKLIVARRAKLOFLSOPFHVAEQFTNTPEAFVQLDITRSFGIVD
GEDPHPEAFTWGVIEEAIVAKKAAMAEN"
complement(3196..4071)
/gene="CC3448"
complement(3196..4071)
/gene="CC3448"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, gamma subunit"
/protein_id="AAK25410.1"
/db_xref="GI:13425166"
/translation="MASIKERNNISVKATOKITRKMOWAAAKIRBSODAESARP
FARRIALSVIANLAGVSDGPKFLAKTGDDRHVVYAADAAGLGAGTTSTVAR
AHIGLIQOGDAIVGWVCCKVTRAKLTPAGSIGVEFPDLSSTRQUTLSVAQPIADY
IRETEAGETDVLTFTFSRFSYOOLITGQQLPAVEEGAASGTAIVEYEPEEA
ITKELETIESAEAV"
complement(4106..5638)
/gene="CC3449"

```

```

CDS      complement(4106..5638)
        /gene="CC3449"
        /note="identified by match to protein family HMM"
        /codon_start=1
        /transl_table=11
        /product="ATP synthase F1, alpha subunit"
        /protein_id="AAK25411.1"
        /db_xref="GI:13425167"
        /translation="MDIRAAEISALIKSOIANGEAASVDSQVLSVSGGSIATYGL
        DNQAGMELEFPKAGVGMALINERDVAVITRQDEIRKEDSEVRNLDGETVDVPR
        GLGRVNPGEPIIDGKPIQYTERNRVDVAPETIDKSVHEPVQGLKSIDLIPV
        GRQREIIGDROTKTAVADITLNRKANAGDESAKLYCVVAIGOKRSYVAQIY
        KTLBEHGALEYITVVVAASAPLQIYLAEPAGCAGMEFNDGLHGLITIDLSKQD
        VAYRQSLRLRRPGREAPGDVYILSRLEERAKINDENGSGSLALPIITQDQD
        VSAIPIYNTISITDGIETDLEFYQIRPAVNVGISVSVGSAQAKAKQVAGPIK
        GEIAYREMAAFKFGSDLDASTOKMKAREERLTLELKQPIYAFLSVEQCVIYAGT
        RGYLDGIPTSVSRPFAEFLARLHSHADLEGIIRKALDKLDELNTLKALDSFST
        FA"
        complement(5691..6245)
        /gene="CC3450"
        complement(5691..6245)
        /gene="CC3450"
        /note="identified by match to protein family HMM"
        /codon_start=1
        /transl_table=11
        /product="ATP synthase F1, delta subunit"
        /protein_id="AAK25412.1"
        /db_xref="GI:13425168"
        /translation="MADERKATPDAGCRYASQSLFELTENGSLQKVEADSLKAMAD
        SADRLIASPFAESADKRGITAVAKKGFQPLTKFGLVANQRTGLDLAISAF
        VLSARHKGVTVAIEVSAALSPQAKQVOTALQALGKTPVSTFVDPSSLGLKVR
        VSRLEFDSLRSRLSLKFKALKRA"
        6338..6742
        /gene="CC3451"
        6338..6742
        /gene="CC3451"
        /note="identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAK25413.1"
        /db_xref="GI:13425169"
        /translation="MDGRRCRAPSEANLITRAFESRNQKATERSLPRSDGISPR
        RPPAPRRRGAVHAPARRSGSWVRPPRCORGRSPRHRSRARRRRLRPPRCRE
        RIPVGRGSCPMVSTSPVAGKEKAAGPAA"
        complement(6831..7283)
        /gene="CC3452"
        complement(6831..7283)
        /gene="CC3452"
        complement(6831..7283)
        /note="identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAK25414.1"
        /db_xref="GI:13425170"
        /translation="WLSDPKRRMGVRRRHPRGRASTRISPIIDLTATPKPEGFIM
        KLRSLAAVAALSVSMFAGSADGRIALALDPAKRVKVVAGVAFVCEGECVSAQ
        APSRALTAAACKALAKEVGRVAAFGGKRSIDAEPLANQMSAKYAAK"
        7328..7864
        /gene="CC3453"
        7328..7864
        /gene="CC3453"
        /note="identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAK25415.1"
        /db_xref="GI:13425171"
        /translation="MRFASGSIADLSPDAPRGMNRRLKPODLDVDPALGIGVQAT
        DHALGAGVAVLADPALDVGOGSRTGEGEAGHOGGHRADGGCHRRDPRDPP
        AEDGIQTIDOKLAVRLILHHPDHDPEDGALDNRGQOHRHQAQNTKSDKQSEKCADRE
        GGHGDHAAVTPFMSRTG"
        7951..9258
        gene

```

```

CDS      /gene="CC3454"
        7951..9258
        /gene="CC3454"
        /note="identified by match to PFAM protein family HMM
        PF00441"
        /codon_start=1
        /transl_table=11
        /product="acyl-CoA dehydrogenase family protein"
        /protein_id="AAK25416.1"
        /db_xref="GI:13425172"

Query Match      5.68; Score 99.2; DB 1; Length 14710;
Best Local Similarity 53.58; Pred. No. 0.014;
Matches 213; Conservative 0; Mismatches 168; Indels 3; Gaps 1;

QY 56 CTTTGACCTGCTGCTGCGGGGACCCCTCATCGACGGAGCAACACCCGGGGGCGC 115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14609 CCTAGCATCTGTGCTCTGCGGAGCGCGCTGTTCGACGGCGCGCGGCGCTTCC 14550
QY 116 GCGCGACCTGGCGGTGCGCGGACCGCATCGCCGCGCATCTGTCGAGCGCG 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14549 AGCGGATGTCGGCGTGGGAGCGCGCATCGTGGGTCGGAAGGCGCTCCCGCG 14490
QY 176 CCGGGACACCCGGGTGACCTGTGCGGCTGTGTGCGCGCGCGCTTCATGATCGC 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14489 GCGCGAGAGA--GATGACGCCGAGAGGCGGATCGTACCCCGGCTGCGATATTC 14433
QY 236 ACACCCAGACGACATACCTGCTCAGGCGTGCAGATAGACGCCAATCTCGGAG 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14432 ACACCCATATGACGCGGACGACGACGATGACGCTGACCTGCGCCACGTCGCGCAG 14373
QY 296 GCGTCACACGCGGTGTCACGGGCAATTCGCGCATACCTGCGCGCGCGCGCACGCCA 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14372 GCGTGACACGCGGTGTATGCGCAATTCGCGCGCGCGCTTCGCGCGCGCGGAGC 14313
QY 356 ACCGCGCGCGCGCGCGCGCGCTGATGACGAGCGCGCTTACCTTTCGAGCGCTTCG 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14312 ATCAGCAGCGCGCTGATCGCGCTGATGGAAGCGTCGAGACATCCCTTCGCCGTGTA 14253
QY 416 CCGACTACTGAGCGCGCTTCGCGG 439
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14252 CCGAGGCGCTGCGCGCTGCGG 14229

RESULT 15
AX211705 LOCUS 65140 bp DNA linear PAT 06-SEP-2001
AX211705 DEFINITION Sequence 1 from Patent WO0159126.
AX211705 ACCESSION AX211705
VERSION AX211705.1 GI:15523937
KEYWORDS Streptomyces noursei.
SOURCE Streptomyces noursei.
ORGANISM Bacteria; Filicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
1 (bases 1 to 65140)
Zotchev,S.B., Sekurova,O.N., Fjærevik,E., Brautaset,T.,
Stroem,A.R., Valia,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
Patent: WO 0159126-A 1 16-AUG-2001;
Norwegian Teknisk Naturvitenskapelige Universitet (NO); STRIETELSEN
IND OG TEKNISK FORSKNING VED NORGENS TEKNISKE HOGSKOLE (NO);
ALPHARMA AS (NO); Sivant AS (NO); Zotchev, Sergey Borisovich
(NO); Sekurova, Olga Nikolayivna (NO); Fjærevik, Espen (NO);
Brautaset, Trygve (NO); Stroem, Arne Reidar (NO); Valia, Svein
(NO)

FEATURES
source Location/Qualifiers
1..65140
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"

```

Search completed: May 11, 2003, 03:03:11  
Job time : 4976 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 11, 2003, 12:05:25 ; Search time 54.5 Seconds

(without alignments)  
6201.993 Million cell updates/sec

Title: US-10-009-782-1  
Sequence: 1 gaatccactgacgcgcga.....ccctgagctacgagaagctt 1758

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=zip  
-O=/cgs2.1/USPTO.spool/US10009782/rmatc.07052003.122517-23148/app-query.fasta.1.1927  
-DB=PIR-73 -OPMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNIT5-bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pic -NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USPR=US10009782.ecgn.1.1.46.gmatc.07052003.122517-23148 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAMP -LARGEORDER -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database: PIR-73:

1: p1r1:  
2: p1r2:  
3: p1r3:  
4: p1r4:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2511	76.1	484	2	JC4394	aminoacylase (EC 3
2	1000.5	30.3	488	2	JC4165	N-acyl-D-glutamate
3	942	28.6	526	2	B75202	d-aminoacylase (as
4	423.5	12.8	1106	2	J00405	hypothetical 119.5
5	407	12.3	581	2	B87678	hypothetical prote
6	338.5	10.3	529	2	T45134	hypothetical prote
7	337	10.2	924	2	S27923	gene lrf3 protein - hu
8	321.5	9.7	660	1	Q08B3	BHLH protein - hu
9	312	9.6	660	1	Q08B3	BHLH protein - hu
10	307	9.3	680	2	S21323	probable endogluca
11	296.5	9.0	1367	1	S48478	glucan 1,4-alpha-g
12	296	9.1	1791	2	T02345	hypothetical prote
13	294	8.9	611	2	D70928	hypothetical prote
14	289.5	8.8	3020	2	A43932	mucin 2 precursor,

C	15	287.5	8.9	1106	2	J00405	hypothetical 119.5
	16	284.5	8.6	1733	1	B45344	probable nuclear a
	17	282.5	8.6	437	2	C39135	hypothetical prote
	18	282.5	8.6	1460	1	EBDETF	immediate-early pr
	19	277.5	8.4	494	2	G84348	hypothetical prote
C	20	274	8.4	1733	1	B45344	probable nuclear a
	21	273.5	8.3	1027	2	S28774	collagen alpha cha
	22	272.5	8.3	1414	1	S23809	collagen alpha 2(I
	23	270.5	8.2	507	2	T44768	antifreeze glycop
	24	270	8.2	1464	2	GCHU15	collagen alpha 1(I
C	25	267.5	8.2	1958	2	B40505	hypothetical prote
	26	267	8.1	825	2	JC4163	DNA-binding prote
	27	265.5	8.0	1791	2	T02345	hypothetical prote
	28	265.5	8.0	3570	2	T45025	mucin MUC5B
	29	264.5	8.0	1466	1	GCHU17	collagen alpha 1(I
	30	264	8.0	1464	2	CS9856	collagen alpha 1(I
	31	263	8.0	1188	2	S49915	collagen alpha 1(I
	32	263	8.0	1453	2	S21626	extensin-like prot
	33	262	7.9	1344	1	A35175	collagen alpha 1(I
	34	262	7.9	1958	2	B40505	mucin 1 precursor,
C	35	261.5	8.1	640	2	T08179	hypothetical prote
	36	261	7.9	779	1	CGBO15	LRG5 protein - chl
	37	261	7.9	839	2	F75518	collagen alpha 1(I
	38	260.5	7.9	580	2	T43481	hypothetical prote
	39	260.5	7.9	964	1	CGCH25	probable mucin DKF
	40	260.5	7.9	1042	1	CGCH15	collagen alpha 2(I
	41	258.5	7.8	2796	2	JC4743	collagen alpha 1(I
	42	255.5	7.7	1151	2	T18535	fatty-acid synthas
C	43	253	7.8	738	2	E87627	high molecular mas
	44	251	7.6	1419	2	A41182	hypothetical prote
	45	251	7.6	1487	2	B41182	collagen alpha 1(I

#### ALIGNMENTS

##### RESULT 1

JC4394  
aminoacylase (EC 3.5.1.14) - Alcaligenes xylosoxydans subsp. xylosoxydans A-6  
N: Alternate names: N-acyl-D-amino acid amidohydrolyase  
C: Species: Alcaligenes xylosoxydans subsp. xylosoxydans A-6  
C: Date: 20-Jan-1996 #sequence=revision 19-Apr-1996 #text=change 13-Sep-1998  
C: Accession: JC4394  
R: Nakayama, M.; Katsuno, Y.; Hayashi, S.; Miyamoto, Y.; Sakai, K.; Moriyuchi, M.  
Biosci. Biotechnol. Biochem. 59, 2115-2119, 1995  
A: Title: Cloning and sequencing of a gene encoding D-aminoacylase from Alcaligenes x  
A: Reference number: JC4394; MIMD:96100942; PMID:8541651  
A: Accession: JC4394  
A: Molecule type: DNA  
A: Residues: 1-484 <NAK>  
C: Comment: This enzyme, which catalyzes the hydrolysis of N-acyl derivatives of neur  
residue of zinc ion or EDTA.  
C: Gene: dan  
C: Superfamily: aminoacylase  
C: Keywords: hydrolase  
F: 68-70/Region: zinc binding

##### Alignment Scores

Pred. No.: 1.64e-128  
Score: 2511.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 76.11%  
DB: 2  
Gaps: 0

US-10-009-782-1 (1-1758) x JC4394 (1-484)

QY 34 ATGTCCCATTCGATTCGACCTGCTGCGGGCGGACCCGATCGAC 93  
DB 1 MetSerGlnSerAspSerGlnProPheAspLeuLeuAlaGlyIleThrLeuIleAsp 20  
QY 94 GGAGACACACCCCGGGGGGGCGGCGGACCTGCGGGCGGCGGACGCGACCGCCGCC 153

```

Db 21 GlySerAsnThrProGlyArgArgAlaAspLeuGlyValArgGlyAspArgIleAlaAla 40
OY 154 ATGGCGATCTGTGAGACCGCGCGGACACACCCGGGTGACGTGTGCGGGCTGTGGTTC 213
Db 41 IleGlyAspLeuSerAspAlaAlaIleThrArgValAspValSerGlyLeuValVal 60
OY 214 GCGCCGGCTTCATGAGTCCGACACCCACGACGACACTACCTCTCAGAGGTGCGGAC 273
Db 61 AlaProGlyPheIleAspSerHisThrHisAspAspAsnIleLeuAspArgArgAsp 80
OY 274 ATGAGCGCCAGATCTGCGAGCGCTCACACAGGTGTGACGGGCAATTTGGCGCATCAGC 333
Db 81 MetThrProGlyIleSerGlnGlyAlaThrThrValAlaIleThrGlyAsnGlyIleSer 100
OY 334 CTGGCGCGCTGGCGGACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 393
Db 101 LeuAlaProLeuAlaIleAlaAspProProAlaProLeuAspLeuAspGlnGly 120
OY 394 TCTTACCGTTCGAGCGCTTGGCGACTACCTGAGCGGTGCGGGCCACCGCGCGCGC 453
Db 121 SerTyrArgPheGlnArgPheAlaAspTyrLeuAspAlaLeuArgAlaThrProAlaAla 140
OY 454 GTCAACGCGCGCTGTATGTGGTGGCATTCACGCTGGCGCGCGCGCGCGCGCGCGCTG 513
Db 141 ValAsnAlaAlaGlyMetValGlyHisSerThrLeuArgAlaAlaValMetProAspLeu 160
OY 514 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 573
Db 161 GlnArgAlaAlaThrAspGlnGlyIleAlaAlaMetArgAspLeuAlaGlnGlyAlaMet 180
OY 574 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 633
Db 181 AlaSerIleAlaIleGlyIleSerThrGlyAlaPheTyrProProAlaAlaArgAlaThr 200
OY 634 ACCGAGAGATCATCGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 693
Db 201 ThrGlnGlnIleIleGlyValGlyArgProLeuSerAlaHisGlyIleIleIleIleIle 220
OY 694 CAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 753
Db 221 HisMetArgAspGlnGlyIleHisIleValAlaAlaLeuGlnGlyIleIleIleIleIle 240
OY 754 CCGGACCTGAGCTGCGCGGTGTGATGTGCGACCAAGATCATGGGCGCGCGCGCGCGC 813
Db 241 ArgGlnLeuAspValProValAlaIleSerHisHisLysValMetGlyIleProAspPhe 260
OY 814 GCGCGCTGCGCGGAGAGCGTGCCTGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 873
Db 261 GlyArgSerArgGlnThrLeuProLeuIleGlnAlaAlaMetAlaArgGlnAspValSer 280
OY 874 CTGAGCGCTATCCCTACGTGCGCGCGCTGCACCATGCTCAACGAGCGCGCGCTGCTG 933
Db 281 LeuAspAlaTyrProTyrAlaIleGlySerThrMetLeuLysGlnAspArgValLeuLeu 300
OY 934 GCGGAGCGACCATCATCATCTGTGTGCAAGCCCTCCCGGAACTAGCGGGCGCGCGC 993
Db 301 AlaGlyArgThrIleIleThrTyrPysProPheProGlnLeuSerGlyArgAspLeu 320
OY 994 GATGAGTGTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1053
Db 321 AspGlnValAlaAlaGlnArgGlyLysSerIleIleIleIleIleIleIleIleIleIle 340
OY 1054 GCGCGCGCGCATCTATCATGATGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGC 1113
Db 341 AlaGlyAlaIleIleIlePheMetAspGlnProAspValGlnArgIleLeuAlaPheGly 360
OY 1114 CCGACCATGATCGGCTCGAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1173
Db 361 ProThrMetIleGlySerAspGlyLeuProHisAspGlnArgProHisProArgLeuTyr 380
OY 1174 GGCACCTTCCCGCGGTGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1233
Db 381 GlyThrPheProArgValLeuGlyHisIleTyrAlaArgAspLeuGlyLeuPheProLeuGln 400

```

```

OY 1234 ACGGCGATGAGAGATGACCGGCTGACCCCGCGCGCTTCGGCGTTCGGCGCGCGG 1293
Db 401 ThrAlaValIlePlysMetThrGlyLeuThrAlaAlaArgPheGlyLeuAlaGlyArgGly 420
OY 1294 CAGTGCAGCGCGGCTACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1353
Db 421 GlnLeuGlnAlaGlyTyrPheAlaAspLeuValAlaPheAspProAlaThrValAlaAsp 440
OY 1354 ACGCGCACCTTCGACACACCTTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1413
Db 441 ThrAlaThrPheGlnHisProThrGlnArgAlaAlaGlyIleHisSerValTyrValAsn 460
OY 1414 GCGCGCGCGTGTGCGCAAGAGCGCGTTCACCGCGCGCGCGCGCGCGCGCGCGCGC 1473
Db 461 GlyAlaProValIleProlGlnGlnAlaPheThrGlyGlnHisAlaGlyArgValLeuAla 480
OY 1474 CGCACGCGCGCGC 1485
Db 481 ArgThrAlaAla 484

RESULT 2
JC4165
N-Acyl-D-glutamate amidohydrolase (EC 3.5.1.-) - Alcaligenes xylosoxydans subsp. x
C:Species: Alcaligenes xylosoxydans subsp. xylosoxydans A-6
C:Date: 12-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 07-May-1999
C:Accession: JC4165
R:Wakayama, M.; Ashika, T.; Miyamoto, Y.; Yoshikawa, T.; Sonoda, Y.; Sakai, K.; Mo
J. Biochem. 118, 204-209, 1995
A:Title: Primary structure of N-acyl-D-glutamate amidohydrolase from Alcaligenes x
A:Reference number: JC4165; MUID:96015170; PMID:8537313
A:Accession: JC4165
A:Molecule type: DNA
A:Residues: 1-488 <MAX>
A:Cross-references: DDBJ:D45918
A:Note: The authors translated the codon CAG for residue 132 as Ala, GGC for resid
C:Comment: This enzyme catalyzes the hydrolysis of N-acyl derivatives of various D
A:Gene: dag
C:Superfamily: aminocyclase
C:Keywords: hydrolase

Alignment Scores:
Pred. No.: 8.09e-47 Length: 488
Score: 1000.50 Matches: 223
Percent Similarity: 59.38 Conservative: 62
Best Local Similarity: 46.46 Mismatches: 190
Query Match: 30.33 Indels: 5
DB: 2 Gaps: 4

US-10-009-782-1 (1-1758) x JC4165 (1-488)
OY 52 CAGCCCTTCGACCTGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 111
Db 3 GlnLysLeuAspLeuValIleGlnGlyIleTyrPvalIleAspGlyLeuGlyArg 22
OY 112 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 171
Db 23 ArgArgAlaAspValGlyIleArgGlyGlnArgIleAlaIleIleGlyAspLeuSerAla 42
OY 172 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 231
Db 43 AlaProAlaAspArgArgLeuAspAlaGlyIleArgIleValAlaProGlyPheIleAsp 62
OY 232 TCGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 291
Db 63 ThrHisGlyHisAspAspLeuMetPheValGlnLysProGlyLeuGlnTyrPheSer 82
OY 292 CAGCGCGTCAACAGGTGTACAGCGCGCAATTTGGCGCGCGCGCGCGCGCGCGCGC 348
Db 83 GlnGlyIleThrSerValValAlaGlyAsnGlyIleSerGlyAlaProAlaProLeu 102
OY 349 CAGCCCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 408

```

[illegible]

Db 461 Phe---ProGlnAlaProProSerHisArgProClyArgIleLeuAlaArgAspAlaSer 479

RESULT 3

B75202

d-aminoacylase (aspartate, glutamate etc) PAB0090 - Pyrococcus abyssi (strain Orsay)

C/Species: Pyrococcus abyssi

C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C/Accession: B75202

R/anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A/Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome

A/Reference number: A75001

A/Accession: B75202

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1526 <K&N>

A/Cross-references: GB:AJ248293; GB:AL096836; NID:95457433; PIDN:CA949065.1; PID:954

A/Experimental source: strain Orsay

C/Genetics:

A/Gene: ndad; PAB0090

C/Superfamily: aminoacylase

Alignment Scores:

Pred. No.:	1,19e-43	Length:	526
Score:	942.00	Matches:	213
Percent Similarity:	55.56%	Conservative:	82
Best Local Similarity:	40.11%	Mismatches:	172
Query Match:	28.55%	Indels:	64
DB:	2	Gaps:	9

US-10-009-782-1 (1-1758) x B75202 (1-526)

OY 58 TTCGACCTGCTGCTCGGGCGGCACCTCATCGACGAGGACACCCGGGGCGGC 117

Db 4 TTTAAATLleValIleLysasnGlyLysIleValAspGlyThGlyAsnProTrpPheArg 23

OY 118 GCCGACCTGGCGTGGCGCGCGACGCATCGCCGCATCGGCAATCTGTGCGACCGCC 177

Db 24 ThrAspIleGlyIleLysAspGlyLysIleValLysIleGlyLysIleLysGlnAspGly 43

OY 178 GCGCACACCGGGGTGCGACGTGCGGGCGTGGCGGCGCGCGCTTATCATCTGCAC 237

Db 44 GlnValTh---IleAspAlaSerAsnLeuIleValAlaProGlyPheIleAspMetHis 62

OY 238 ACCCAGCAGACACTACCTGCTCAGCGGTGCGACATGACGCCCAATATCGCAGGAG 297

Db 63 AlaHisAspAspLeuIlePhePheLysAspArgPheAsnArgAlaLysLeuLeuIndGly 82

OY 298 GTACACAGCGTGTACACGGCAATGCGGCATCACCCTGCGCGCGTGGCGGCACGCCAAC 357

Db 83 ValThrThValValSerGlyAsnGlyIleSerValAlaProValAsnGlu----- 100

OY 358 CCGCGCGCGCGCGTGGACCTGCTGCGACGAGAAGCGCGCTTAC----- 399

Db 101 -----GlnMetLeuAspValLeuLysSerTrpValGlyIleLeuGlyLys 115

OY 400 -----CGTTTGAGGCGCTTCCGCCACACTGACGCGTTCGGGGCCACG--- 444

Db 116 GluValGluPheLysTrpArgSerTrpGlyIlePheLeuAspAlaLeuGluGluValGly 135

OY 445 CCGCGCGCGCGCAGCGCGCGCTGATGGTGGCCATTCACAGCGCGCGCGCGCGCATG 504

Db 136 ProLeuGlyThrAsnPheValGlyLeuValAlaLysIleGlyThrAsnArgIleValAlaMet 155

OY 505 CCGGACTTGACGCGCGCGCCACCGACGAGAGAATTCGCGGCATCGCGGAGCGCGCAG 564

Db 156 GlyMetGluAlaArgAspProThrGluGluGluLeuGlyArgMetLysGluLeuLeuAla 175

OY 565 GAACCGCATGGCAGCGCGCGCATCGGCATTTGCACGCGCGCGCTTACACCGCGCGCGCC 624

Db 176 LysSerMetGluGluGlyAlaPheGlyMetSerSerGlyLeuIleTrpProGlyVal 195

OY 625 CGCGCCACACCGAAGATCATGAGAGTGTGCGCGCGCGTGAACGCGGCTATGCGCGATC 694

#### RESULT 4

hypocretin-like 119.5K protein (vira region) - *Micrococcus luteus*  
N; Alternate names: ORF 1 protein  
C; Species: *Micrococcus luteus*

C:\Data: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 04-Feb-2000

A:Molecule type: DNA  
A:Residues: 1-1106 <SHIT>  
A:Cross-references: EMBL:X15867  
A:Note: this reading frame extends between two stop codons and does not begin with a start codon  
A:Note: the gene encoding this protein overlaps uvra gene  
C:Superfamily: collagen alpha 1(I) chain: fibrillar collagen subunit 1

Alignment scores:

Pred. No.:	1.4e-15	Length:	1106
Score:	432.50		

Percent Similarity:	423.50	Matches:	239
Percent Identity:	32.848	Conservative:	28

Best Local Similarity:	29.408	Mismatches:	252
Query Match:	12.848	Indels:	294

DB:	2	234
Gaps:	2	47

US-10-009-782-1 (1-1758) x J00405 (1-1106)

38 CCCAATCCGATTCCACGCGCTTGACCTGCTCGCGGGCGGCACCCCTCATCGAGGCA 97

Db 42 ProValSerCysProCysProProLeuCysTTPProArgAla-----Clnclosure 50

[illegible][illegible][illegible][illegible]

8 GARGARGALALEUGLYARGALAGLYPRONHISARGARGPRGGLYARGPRGALAGL 98  
 98

206 TGGTGGTCGGCCGGCTTCATCGACTCGCACACCCACGACGACAACTACTGCTCAGGC 265

Db 98 oGIUGIYArgGly-----CysGlnLeuPro-----Al 107

QY 266 GTCGACATGACGCCCAAGATCTCGCAGGGCGTCACCAACGGTGGT-----CAGCCCA 319

Db 107 aclyvArqHis-----G|vArvValu|ecl"i"fo"e"v"e"l"-----17 100

320 A-----TTGCCCCAAT-----GTGCCCTGGTGGT  
cagtagcaatnrsolyseivalairgleuglygl 121

121 nfa1fou2201-7 711  
:: || || ||:: ||  
-CAGCCTGGGCGCGCTGGCGCACGCCAACCGGC 361  
2100000001

421 uvalenurfglyleuarghlsaspleuargargglyproalaaleuargvalal 141

362 C-----CGCCCC-----CCCTGGACCTGCTGGACGAGAGCGGCTCTTACCGTTTCG 406

141 aLeuLeuAlaGrTronIaValProGlyProGlyGlyGlnAlaGly-----Ar 157

407 AGCGCTGCGCGA-----CTACTGAGCGCGTT----- 434

157 GGLYLEUHLARGGLYPROVALPROGLYARGVALHLSATGPRGluVAHLSGIPROG 177

435 -----GCGGGCCACGCCCGGGCGGCGGTCAACGC-----

```

177 APPROXIMATE[SSerG] vaaenycwycv... 461
      |||:::|||||
db

```

V : 462 -----  
.....nphnnsnsgaspreuatglenhsalaproalaucgl 197

107 -----CGCCTGTATGGTGGCCATTCAC---GC 487

157 YINCYSAAGGLYALAALAEUPROALAVALARGALAGLIGLUPROALAASPROAL 217

488 TCGCGCGCGGTCATGCCGACTTGACAGCGCGCCGAC-----CG 529

217 aaLaaspArgGlyProaLaArgGlyAlaAlaLaArgAlaHisProLeuProGlyAlaArgAl 237

Y 530 ACGAGGAATCGCGGCCATGCG-----GGACCTGGCCGAGGA-----AGGCATGG 574

237 aArgGlyProArgProGlnGlyArgValArgGlyProValGlnGlyProValHisAlaIleI 357

575 CCAAGCGGCGCCATCGGCATTTCGAACGGGGCTTGTAA--CCCCCCCCCCCCCC







US-10-009-782-1 (1-1758) x T45134 (1-529)

```

OY 32 CCAATGCCAATCCGATCCCAAGCCTTC---ACCTGCTGCTGCGG-----76
Db 1 ProAlaProSerAlaThrArgSerProSerTrpSerIleCysSerArgIleProThr 20
OY 77 -----GGGCAACCTCATCG 91
Db 21 SerProAlaThrThrSerSerThrCysArgThrAlaProAlaAlaSerAlaProAlaSer 40
OY 92 ACGGCAACAAACACCCCGGCGCGCGCGCACTGGCGCTGCG-----134
Db 41 ThrAlaGluProHisAlaGluProAla-LysProGluArgAlaGluProArgGluThrG1 60
OY 135 -----CGGCAACCGCATCGCCGCAATCGGCGCATCTGTGCGAGCGCGCGCG 181
Db 60 yThrIleGluProArgGlySerHisTrpArgHisArgArgProAlaGluArgGluIva 80
OY 182 ACAACCGGCTGCACTGCTGCGGCTGCTGCTGCGCGCGCTTCATGCACTGCAACCC 241
Db 80 HisProLeuGluArgPro-----HisGluGluProGluAl 92
OY 242 ACGACACAACTACTCTGCTCAGCGCTGCGCATGACGCCAAGATCTCGAGGCGCTCA 301
Db 92 aArgGlyGluLeuProValArgAspArgAlaGlu-----104
OY 302 CCAACGGTGTACAGGGCAATGCGGCATGACCTGCGCGCGCTGCGCGCACCCCGC 361
Db 105 -----ArgGlyArgArgGluProAl 111
OY 362 CCGCCCCCTGGAACCTGCTGAGACGAGCGGCTTACCCTTCAGAGCGCTGCGCGAC 421
Db 111 aArgProAlaAlaArgAspAlaArgArgHisLeu-----122
OY 422 ACTGAGCGCTGCGGCGCGCGCGCGCGCTCAACCGCGCTGATGCTGCGCAT 481
Db 123 -----ArgLeuGlyAlaHis 127
OY 482 CAACGCTGCGCGCGCGCGCTCATGCGGACTGCAAGCGCGCGCGCGCAAGAAATG 541
Db 128 ----ProAlaArgArgGly-----ValLeuArgArgHisArgArg---HisArg 141
OY 542 CCGGCATCGCGGCACTGCGCGCAAGAACCATGCGCGCGCATGCGCATTCGACAT 601
Db 141 gAlaArgGlyGluArgGlyArgGly-----ProArg 151
OY 602 GCGGCTTCAACCGCGCGCGCGCGCGCGCGCACCAACGAGATCATCGAGTGTGCGCG 661
Db 151 gGluGlnValProArgGlnHisProArg---GlyArgArgAspArgAlaGlyArgProG1 170
OY 662 CGCTGAGCGCGCATGCGCGCATCTACGCCAACCCATGCGCGCGCAAGCGCGACATG 721
Db 170 yLeu-----HisArgArgArgArgArgArgArgArg 180
OY 722 TGGCGCGCTGAGAGAACTT-----CGGCATCG 751
Db 180 gArgGlnGluArgProGlnValArgHisGlyAspArgGlnHisArgAlaAspArgArg 200
OY 752 GCGCGGAGCTGAGCTGCGGCTGATCTGCAACCAAGATCTATGCGCGCAAT 811
Db 200 gProArgAspProArgGlnHisProLeuArgGluGluGlnGlnGlnGlnGlnGlnGln 220
OY 812 TCGCGCGCTGCGGAGAGCGCTGCGCGCATGAGCGCGCATGCGCGCGCGCGAGCTCT 871
Db 220 aArgProLeuArgGly-----ArgGlyGlyGlnGlnGlyProGlyGlyArg 235
OY 872 CGCTGAGCGGCTATCCTACCTGCG-----CGGCTCCACCATGCTCAAGAGAGCGCG 925
Db 235 oAlaGlyArgHisProAlaGluGlyAspArgProArgArgAlaArgGlnGlnGlnGln 255
OY 926 TCGTGTGCGCGCGAGCAACATCATCACTGCTGCAAGCCCTTCCCGAACTGAGCGGCG 985

```

```

Db 255 gAlaAlaHisArgGluAlaLeuHisLeuArgGluArg-----268
OY 986 GCGACCTGATGAAGTGGCGCGCGCGCGCGCGCAATCCAAATACAGACCTGCGCGAGC 1045
Db 269 -----ArgArgGlyArgGly-----AspIleCysArgAl 278
OY 1046 TGCAGCGCGCGCGCGCATCTTCAATGATGACGAAACCGAGTGCAGCGCATCTGCG 1105
Db 278 aGlnGlyArgAlaArg-----ArgAlaArgArgAlaGlyGlyGly 292
OY 1106 CGTTCGCGCGCGCGCATATGCTGCGCA-----CGGCTCGC 1141
Db 292 yLeuProArgArgGlyuAspArgValArgAlaHisArgProArgProArgGlyGly 312
OY 1142 CGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1201
Db 312 yAlaAlaArgLeuAspArgAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 332
OY 1202 A---TGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1232
Db 332 uArgHisProArgProAlaAspArgLeuProHisGlyGlyProGlnGlyValAlaArgLeuAs 352
OY 1233 ---GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1288
Db 352 rHisProGlnGlyLeuGlnGlyGlyAlaAlaGlySerArgArgHisProHisArgLeuArgAl 372
OY 1289 GCGGCGAGCTGACGCGCGG---GTACTTGGCGCGCG-----1322
Db 372 aArgLeuHisGlnGlyArgGlyAspLeuArgGlyArgProArgArgArgArgArgArg 392
OY 1323 -GGTGTGTTGACCGCGG-----CAAGTGGCGCGCGCGCGCGCGCGCGCGCGCG 1372
Db 392 gArgGlyProArgGlnGlyGlnGlyAlaHisGlyGly-----GlnGlyLeuArgHisAla 410
OY 1373 CTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1432
Db 410 aGlyArgArgArgArgGlyGlyAlaProSerGlnLeuAlaValLeuArgGlnGlyGlyAla 430
OY 1433 AGCAGCGCTTCAAC-----GGCGACCATGCGCGCGCGCGCGCGCGCGCGCGCG 1483
Db 430 nAlaSerAla-ThrLeuAlaLeuTrpMetThrSerGlyArgLeuArgGlyGlyAlaG 450
OY 1484 CCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1543
Db 450 yLeuLeuCysArgAspSerValGluSer-----459
OY 1544 CAACCTGAGCAACCAACCGCTCATGCGCCCTCCCTCC---GCTGCAATACGCGCGCAC 1600
Db 460 --HisLeuAspValThrThrTrpArgAlaAspAlaSerIleSerArgArgThr 476
OY 1601 CCGATATCGTGGCAAGAAATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1656
Db 477 -----TrpThrValMetAsnThrArgLeuArgGlnAlaGlnIleuSerLeuA 492
OY 1657 -----CGGAAATGACCTGCAAGACCTGTCGACGCG 1689
Db 492 lAlaAlaArgArgProThrThrSerSerIleSerArgSer 505

```

RESULT 7  
S27923  
gene Lf3 protein - human herpesvirus 4  
C/Species: human herpesvirus 4, Epstein-Barr virus  
C/Date: 19-Mar-1997 #sequence-revision 19-Mar-1997 #text-change 31-Jan-2000  
C/Accession: S27923  
R/Parker, B.D.; Bankier, A.; Satchwell, S.; Barrell, B.; Farrell, P.J.  
submitted to the EMBL Data Library, August 1990  
A/Description: Sequence and transcription of Raji Epstein-Barr virus DNA spanning the  
A/Accession: S27923  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-924 <PAR>  
A/Cross-references: EMBL: M35547; NID: g330420; PIDN: AAA45896.1; PID: g330421

C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

## Alignment Scores:

Score: 6.58e-11 Length: 924  
Percent Similarity: 33.700 Matches: 210  
Best Local Similarity: 34.588 Conservative: 30  
Query Match: 30.264 Mismatches: 287  
DB: 10.228 Indels: 168  
Gaps: 31

US-10-009-782-1 (1-1758) x S27923 (1-924)

```

QY 45 CSATTCACGAGCCCTTCGACCT--GCTGCTCCGCGCGCGCCACCTTCATCGACGCGCA 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 ArgGlyProCysLeuAlaProGlySerGlyLeuGlyAlaHisProHisProArgSer 25
QY 102 -----CACCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 26 GlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProAlaProGly--- 44
QY 132 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 45 ---ProGluProArgThrArgLeuGlnProAlaThrProArgSerGlyAlaAlaAsp 63
QY 180 -----GCACACCCGCGGT-----CGAGCTGCGCGCGCGCGCGCGCGCGCGCG 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 ProAlaAspProValGlyHisProAlaAlaProAlaProGlyProGluProArgThr 83
QY 219 CGGCTTCATCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 ArgLeu---GlnProAlaThrProArgSerGlyAlaHisProAlaAspProVal 102
QY 279 GCCAGATCTCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103 GlyHisProAlaAlaProAlaProGlyProGluProArgThrArgLeuGlnProAla 122
QY 339 GCC-----GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 ThrProArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaPro 142
QY 369 -----CCTGACCTGCTGACGAGCGGCTC-----TTACGCTTTCAGCGCTT 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 ArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThrProArgSerGly 162
QY 414 CGCGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 AlaAlaAspProAlaAspProValGlyHis-ProAlaAlaProAlaGly-ProGlyProG 182
QY 474 GCGCGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 LurProArgThrArgLeuGlnProAlaThrProArgSerGlyAlaAlaAspPro--- 200
QY 534 GGAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 ---AlaAspProValGly-----HisProAlaAlaProAlaGlyAlaP 213
QY 594 TTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 213 ToglProGluProArgThrArgLeuGlnProAlaThrPro-ArgSerGlyAlaAla 232
QY 642 -----GATCATCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 AspProAlaAspProValGlyHisProAlaAlaProAlaGlyAlaP 250
QY 693 CCACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 ProArgThrArgLeuGlnProAlaThrProArgSerGlyAlaAlaAspProAlaAsp 270
QY 753 CCGCGAGCTGAGCTGCGCGGTGATCTGCGACACACAGGTCATGGGCGCA 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 Pro-----ValGlyHisProAlaAlaProArgThrProGluProArg 286
QY 804 GCCCAATTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 848

```

```

Db 287 ThrArgLeuGlnProAlaThrProArgSerGlyAlaAlaAspProAlaAspProVal 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 849 CCGCATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 GlyHisProAlaAlaProAlaProGlyProGluProArgThrArgLeuGln-Pro 326
QY 900 CTCACCATGCTCAAGCAGACCGCG-----TGCTGCTGCGCGCGCGCGCGCG 944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 326 LurProArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaP 346
QY 945 CATATCATCTGGTGCAGAACCTTCGCCGAATGAGCGCGCGCGCGCGCGCGCGCG 1004
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 RohArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThr-----ProArg 364
QY 1005 GCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 rgsSerGlyAlaAlaAspProAlaAspPro-ValGlyHisProAlaAlaProAlaP 383
QY 1059 -----CGCATCTACTTCATGATGAGACCGCGCGCGCGCGCGCGCGCGCG 1106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 GlyProGluProArgThrArgLeuGlnProAlaThrProArgSer-----Gly 400
QY 1107 GTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 AlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgThrProGlu 420
QY 1167 CTTGTGGCGCACCTT----- 1181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 ProArgThrArgLeuGlnProAlaThrProArgSerGlyAlaAlaAspProAlaAsp 440
QY 1182 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 ProValGlyHisProAlaAlaProAlaProGlyProGluPro----- 455
QY 1242 ATGAGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 -----ArgThrArgLeuGlnProAlaThrProArgSerGly 468
QY 1302 GCGCGGCTACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 AlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgThrProGlu 488
QY 1361 CTTTGACACCTTA----- 1375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 489 ProArgThrArgLeuGlnProAlaThrProArgSerGlyAlaAlaAspProAlaAsp 508
QY 1376 -----CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 509 ProValGlyHisProAlaAlaProArgThrProGlyProGluProArgThrArgLeuGln 528
QY 1424 TCTGCGAAGACGAGCGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 ProAlaThrProArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAla 548
QY 1466 TGCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 549 AlaProArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThrPro-Arg 568
QY 1520 ACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 568 gSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgThr 587
QY 1565 ACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 588 ---GlyProGluProArgThrArgLeuGlnProAlaThrProArgSerGlyAlaAla 606
QY 1625 T-----GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 AspProAlaAspProValGlyHisProAlaAlaProArgThrProGluProArg 626
QY 1665 -----GACCTGCAAGACCTGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 626 gTtAgtGleuclnPrOalAatThrProArgArgSerGlyAlaAlaAspPrOalAspProVa 646

OY 1715 CCAAGCCGAGACTGGCCCAATGGCCCT 1742

Db 646 lGlyHisProAlaAlaIleProAlaGlyAlaPro 655

**RESULT 8**

C:Species: human herpesvirus 4, Epstein-Barr virus  
 C:Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 23-Aug-1997  
 C:Accession: A03742  
 R:Bankier, A.T.; Dellingner, P.L.; Farrell, P.J.; Barrell, B.G.  
 Mol. Biol. Med. 1, 21-45, 1983

Alignment Scores:	
Pred. No.:	4 4e-10
Score:	321.50
Percent Similarity:	31.638
Best Local Similarity:	29.068
Query Match:	9.738
DB:	1
Length:	660
Matches:	187
Conservative:	18
Mismatches:	256
Indels:	183
Gaps:	29

US-10-009-782-1 (1-1758) x QDBEZ (1-660)

39 CGATTCGATTCACGAGCCCTTCGACCTGCT----- 68  
 113 ProGlnArgSerProAlaGlyThrArgGlnAlaGlyTyrAlaLeuGlyGlnGlySerAlaGly 132  
 69 -----GCTGGCGGGCGGACCCCTCATCGAGCGGACCAACCCCGGGGCGGCGCGCA 122  
 133 LeuGlySerArgGlyProAlaGlyProHisProAlaPheGlnValGlnTrpSerAlaArgAsn 152  
 123 CTTGGCGGCTCGGGCGGACGCGATCGCGCGCATGTCGTGGAGACCGCGCGCA 182  
 153 ProGlyCysPro-----ArgThrTyrAlaArgAlaGlySerGlyAlaGlnArgGly 168  
 183 CACCGGGGTGACGCTGTGGGCTGTGGTCGCGCCGCGCTTCATCGACTCGCACACCA 242  
 169 HisPro----- 170  
 243 CGACGACAATCACTGCTCAGGCGTGGCAGATGACGCCCAAGATCTGCAGGCGCTAC 302  
 171 -----ProProGlyAlaGlyGlnArgProSerGlyProThrGlyAlaArg--- 185  
 303 CACGGTGTACAGGGCAATTGGCGCATCGACCTGGCGCGCGTGGCGGACGCCAACCCCGCC 362  
 186 -----ProAlaAlaProGlyAlaProGlyThrPro 195  
 363 CGCCCCCTGGAAGCTGTCGAGCAGAGCGGCGCTCTTACCGTTTCAGAGGCTTCGGCGACTA 422  
 196 AlaAlaProGlyProGlyGly--GlyAlaAlaValPro--SerGlyAlaThrProHisP 214  
 423 CCTGGAACGCTTGGGGCA-----CGCC 446  
 214 roGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgG 234  
 447 GCGGCGCTCAACGCCCTGTATGTGGCCATTCACGCTTCGCGCGCGCGCGGCGATGCC 506  
 234 InGluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysPro----- 249

OY	507	GGACTGTGAGGGGGCCGCCACCGACAGAAATCGCGGCATCGGAGACTGGCCGAGA	567
Db	250	-----AlaGlyProProProThiAlaSerGly-----	250
OY	567	AGCCATGCGCAGCGGCGCATCGGCATTTGAGCGCGGCTTACCGCGCCGCGCGCG	622
Db	259	-----AlaAlaAlaGlnAlaGlyThiHisArg-----	271
OY	627	CGCCACACACGAAAGATCATCATAGGTGTGGCGCGCGCATCGGCATCGGCATCTA	686
Db	271	yspPro-----ArgSerAlaArg--AsnProGlyCysProAlaGlyThiParg-----	285
OY	687	CGCCACCCACATGCGCGACAGAGGAGGACATCGTGGCGCGC-----GCTGAGAGA	737
Db	286	-----ArgArgSerGlyAlaGlnAlaGlyGlnHisProProProGlyAlaGly	300
OY	738	AACCTTCCGATGCGCGCGGAGCTGGACCTGGCGGTGTGTGATCTCGACACAGAGCAT	797
Db	301	GlnArgProSerGlyProThiArgGlyAlaGArgProAlaAlaProGlyAlaPro-----	317
OY	798	GGCGCAGCCCAATTGCGCGGCTCGCGGAGACGCTGCCGTATCGAGCGCCCATGCG	857
Db	318	-----GlyThiProAla	321
OY	858	GGCGCAGAGAGCTCTCGCTGGACGCGTATCCCTACCTGCGCGGCTCCACACATGTCAGCA	917
Db	322	AlaProGlyProGlyGlyAlaAlaAlaValProSerGlyAlaThiProHisProGlnArg	341
OY	918	GGA-----CCGCGCTGCTGCTGCGCGACGACCATCATCATCTGTGGTCA	962
Db	342	GlySerGlyProAlaAsnProProAlaAlaAlaArgLeuProProGlnArgGlnPro	361
OY	963	GGCCTTCCCGAGTACAGCGGCGCGACCTGATGAAGATCGCGCGAGCGCGGCAATC	102
Db	362	ArgLeuProGlnAlaLeuAlaAlaAla-----GlnArgCysProAlaGlyProPro	378
OY	1023	CAAGTACAGAGCTGTGCGCGGACGCTGACGCGCGCGCGCATCTACTTCATGATGACGA	108
Db	379	ProThiArgSerGly-----AlaAlaAlaGlnArgThiHisArgArgProProGlyCys	396
OY	1083	ACCGCAGCTGACGCGCATCTGTGGTGTGGCGCGACCATGATGAGGCTCGCAGCGCTGCC	1144
Db	397	ProArgSerAlaArgAsnProGlyCys---ProArgThiThrPargArgSerGlyAla	415
OY	1143	GCACGACGAGCGCC-----GCATCGCGCGCTGTGGGCGACCTT	1183
Db	416	GlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGlyProThiArgGlyArg	435
OY	1182	CCCGCGGCTGTGGGCG-----CTATGGCGCGCATCTGGGCT-----	1220
Db	436	ProAlaAlaProGlyAlaProGlyThiProAlaAlaProGlyProGlyGlyAlaAla	455
OY	1221	GTTCCCGCTGA-----GACGCGGATGGAAGATGACGCGCGCTAC	1262
Db	456	ValProSerGlyAlaThrProHisProGlnArgGlySerGlyProAlaAsnProProAla	475
OY	1263	CGCGCGCGCTGTGGCGCGCGCGCGCGG-----GCAGCTGCA	1301
Db	476	AlaAlaArgLeuProProGlnArgGlnGlyProArgLeuProGlnAsnLeuAlaAla	495
OY	1302	GGCGGCTACTTGGCGCGACCTGTGTGTGTGACCGCGCGACAGCTGGCGAGTACCGCAC	1361
Db	496	GlnArgCysProAlaGlyProProProThiArgSerGlyAlaAlaAlaGlnArgThiHis	515
OY	1362	CTTGGACACCCACGAGCGCGC-----CGCGCGCATCATTCCTGTACGTCAAG-----G	1415
Db	516	ArgArgProProGlyCysProArgSerAlaArgAsnPro--GlyCysProArgThiThrP	535
OY	1416	CGCGCGCGTGTGACAGACAGCGCGCTTACCGCGCGACGATCGCGCGCGGCTCGGACG	1475
Db	535	rgrArgArgSerGlyAlaGlnArgGlyHisPro-----	545
OY	1476	CAGCGCGCTGAGCCCGCGCGCGCGCTTACATCCGCGCGTGAACGCGCGGCTGCGC	1535

Db 546 -----ProProGlyAlaGlyGlnArgProSerGlyProThr-----GlyGlyArgProAla 563  
 QY 1536 CCCCCTCCCAACCTGGAGCAACCGCTCAATGGCC-----CTCCCTCCG 1583  
 Db 563 LabProGlyAlaProGlyThrProAlaAlaProGlyGlyGlyAlaAlaValPro 583  
 QY 1584 TCCCAATACGGCCCAACCGCATCTGGGCAAGAGATGCGCGCCGCTGGCCG 1643  
 Db 583 GcGlyAlaThrProHisPro-GlnArgGlySerGlyProAlaAspProAlaAla 602  
 QY 1644 CGAGCCCAAGCCCGCAATATGACCTTCAGACCTP-----GTCCAGCCAGCGG 1694  
 Db 603 ArgLeuProProGlyLysGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArg 622  
 QY 1695 ---CATCCGCTCTCGACCGCTGTCACAGCCGAGCTGGCCAGATCCGCTGAGCTACA 1751  
 Db 623 CysProAlaGlyProProProThrProHisSerGlyAlaAlaAlaGlnArgThrHisArg 642

## RESULT 9

OOBES

BHFL1 protein - human herpesvirus 4 (strain B95-8)  
 C/Species: human herpesvirus 4, Epstein-Barr virus  
 C/Date: 25-Feb-1985 #sequence, revision 25-Feb-1985 #text, change 23-Aug-1997  
 C/Accession: A03742  
 R/Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.  
 Mol. Biol. Med. 1, 21-45, 1983  
 A/Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus  
 A/Reference number: A93065; MUID:85055713; PMID:6092825  
 A/Accession: A03742  
 A/Molecule type: DNA  
 A/Residues: 1-660 <BANK>  
 R/Barer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H  
 Nature 310, 207-211, 1984  
 A/Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
 A/Contents: Annotation: protein coding region  
 C/Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52  
 C/Superfamily: human herpesvirus 4 BHFL1 protein

## Alignment Scores:

Pred. No.: 1,44e-09 Length: 660  
 Score: 312.00 Matches: 181  
 Percent Similarity: 31.598 Conservative: 30  
 Best Local Similarity: 27.108 Mismatches: 232  
 Query Match: 9.628 Indels: 225  
 DB: 1 Gaps: 31

US-10-009-782-1 (1-1758) x OOBES (1-660)

QY 1753 TCTCGAGCTCAGCGCATCTGCCACGT-----CGGCT 1718  
 Db 87 SerAlaGlnSerAlaArgThrGlyProAlaGlnAlaAspHisAlaHisSerAspPro 106  
 QY 1717 TGGACAGGCTGAGACCCGATGCCGCGCTGGACAGCTCTTGACGGGTCAATTTC 1658  
 Db 107 ThrGlyGlyCysSerAspProGlnArgSerProArgThrArgGlnAlaGlyTyrAlaLeu 126  
 QY 1657 GGGCTTCGCTCGCGCGGAGCGCGCGCCATCACTCTCTGCGCCACGATCGGGTG 1598  
 Db 127 GlyGlnGlySerAlaGlyLeuGlySerArgGlyProArgProHisProAlaPheGlnVal 146  
 QY 1597 -----GGCGCTATTGC----- 1586  
 Db 147 GlnTrpSerAlaArgAsnProGlyCysProArgThrTrpArgArgArgSerGlyAlaGln 166  
 QY 1585 -----GAGCGAGGAGGAGGCGCCATGATACGGGT----- 1559  
 Db 167 ArgGlyHisProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgPro 186  
 QY 1558 -----TTGCGTCCAGGCGTGGAGGGGCG----- 1535  
 Db 187 AlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaVal 206

QY 1534 -----GGACGCCCGCCCGTTACGCCGATGTGAAGCGTGGCGCGGCTCAGCGG 1481  
 Db 207 ProSerGlyAlaThrProHisProGlyArg-----GlySerGlyProAlaAsp 222  
 QY 1480 CCGTGGCTCGGACGACG----- 1457  
 Db 223 ProProAlaAlaAlaArgLeuProProGlnArgGlnGluProArgLeuProGlnAspLeu 242  
 QY 1456 CATGCTGCGCGGTGAACCTCTCTTCCAGACCGGG----- 1418  
 Db 243 AlaAlaAlaGlnArgCysProAlaGlyProProThrArgSerGlyAlaAlaGln 262  
 QY 1417 CGCGCTTACGTACACGGAATGATCCGCGCGCGCGCTGGTAGGCT-----GTTCGA 1364  
 Db 263 ArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysArg 282  
 QY 1363 AGTGGCGGTATGG----- 1328  
 Db 283 ThrTrpArgArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlnArg 302  
 QY 1327 CCACCAAGTGGCGGACGATACCGCGCTGACGCT-----GCCGCGCGCG 1283  
 Db 303 ProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAla 322  
 QY 1282 CCAGCGCGAAGCGCGCGCGGTCACTTCATCCATCCAGCGCGCTCCAGCGGGA 1223  
 Db 323 ProGlyProGlyGlyGlyAlaAlaAlaProSerGlyAlaThrProHisProGlyAlaGly 342  
 QY 1222 ACAGGCCAAGTGGCGCGCATAGTGC----- 1193  
 Db 343 SerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlyLysGlnGluProArg 362  
 QY 1192 -----GCACCGCGGAGAGTGGCCACAGCGCGGATCGCGCGCTCGT 1148  
 Db 363 LeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArgSer 382  
 QY 1147 CGTGGCGAGCGCGCGCGGCGGATCATGTGCGGCGGACCGACAGATGCGCTGCACGT 1088  
 Db 383 GlyAlaAlaAlaGlnArgThrHisArg-----ArgProProGlyCysProArgSer 399  
 QY 1087 CGGTTGCTCATGATGAAGTGA-----TGGCGCGCGCGCGCGCTCGGACCGACGT 1031  
 Db 400 AlaArgAsnProGlyCysProArgThrTrpArgArgSerGlyAlaGlnArgGlyHis 419  
 QY 1030 GGTACTGGATTTGCCGCGCTGGCGGCGGACTTCATCCAGTCCGCGCGCGGTGCG 971  
 Db 420 ProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaPro 439  
 QY 970 GGAAGGCTTGCACCAAGTGAATGATGCGCTCGCGCGACGACGCGGTCTGCT--- 914  
 Db 440 GlyAlaProGlyThr-----ProAlaAlaProGlyProGlyGly 452  
 QY 913 -----TGAGCATGGTGAAGCGCGGACGATGAGGATATGCGCTCCAGCGAGA 869  
 Db 453 GlyAlaAlaValProSerGlyAlaThrProHisProGlyLysGlnGluProArgLeuProGlnAsp 472  
 QY 868 GATGCTGGCGCG----- 836  
 Db 473 ProProAlaAlaAlaArgLeuProProGlyLysGlnGluProArgLeuProGlnAspLeu 492  
 QY 835 GCAAGCTCTGCGCGAGCGCGGAAATGGGCTGGCCCATGACCTGTGTGGTGAGATCA 776  
 Db 493 AlaAlaAlaGlnArg-----CysProAlaGlyPro 502  
 QY 775 CCACGCGACGT-----CCAGTCCGCGCGGATGCGGAAGTTCTCTCCAGCGCGCA 722  
 Db 503 ProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysPro 522  
 QY 721 CGATGTCGCGCTCTGCGCGCATGTGGTGGCTAGATGCCCATGCGCGGTCAAGG 662  
 Db 523 ArgSerAlaArg-----Asn 527

[illegible]





Db 760 SerSerAlaPro---ValProThrProSerSerSerThrThrcGluserSerSerAlaPro 778

QY 1391 GCATCCATTCGGCTAGCTAGCTAACGGCCGCCGGCTCTGGCAAGACAGCGCCCTACCGGCC 1450

Db 779 ValProThrProSerSerSerSerThrThrcGluserSerSerAlaProValProThrProSer 798

QY 1451 AGCATGCCGGCCGGCTGCTCGACGCACGCACGGCCGCTGAGACCCGGGGCCAGCCCTACAT 1510

Db 799 SerSerThrThrcGluserSerValAlaProValProThrProSerSerSer---Asn 817

QY 1511 CCGGCGTGAACGGGGCGCGCTGCCGCCCTCCCAACCTGGACGCMAACCGCTACATGG 1570

Db 818 IleThrSerSerAlaProSerSerThrProSerSerSerThrGluserSerSerVal 837

QY 1571 CCCCTCCCT 1579

Db 838 ProValPro 840

RESULT 12

T02345

hypothetical protein KIAA0324 - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 05-Nov-1999

C/Accession: T02345

R/Ricks, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, J.; White, S.; Deng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L. submitted to the EMBL Data Library, March 1998

A/Description: Sequencing of human chromosome 16p13.3.

A/Reference number: 214664

A/Accession: T02345

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1791 <RIC>

A/Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AC08453.1; PID:g2996650

C/Genetics:

A/Map position: 16

A/Introns: 1610/2; 1706/2

A/Note: KIAA0324

Alignment Scores:

Pred. No.:	1,15e-08	Length:	1791
Score:	296.00	Matches:	179
Percent Similarity:	35.61%	Conservative:	50
Best Local Similarity:	27.84%	Mismatches:	221
Query Match:	9.13%	Indels:	194
		Gaps:	27

US-10-009-782-1 (1-1758) x T02345 (1-1791)

QY 1747 AGCTACGGCCATCTGGCCCAAGCTCGGCTTGACAGAGGTGAGACCG----- 1700

Db 619 SerLySerArgLuserProArgArgSerArgSerArgSerArgSerProGluValLysAsp 638

QY 1699 CGATCCCTGCTGGCTCGGACAGGCTTGACAGGCGATTTCCGGCCCTTGCCTGGCGGC 1640

Db 639 LysProArgAlaAlaProArgAlaGlnSerArgSerArgSerPro----- 654

QY 1639 GCAGCGCGCGCCCATCATCTCTTGCCCAAGATATCGGGGTGGGCGCGATATGCGAGCGG 1580

Db 655 -----GluProLysAlaProAlaPro-----ArgAlaLeuProArgArgSerArg 669

QY 1579 AGGAGAGGGCCATGATAGCGTTTGCTTCCAGGGGTGGAGGGGGCGGACG----- 1529

Db 670 SerGlySerSerSerLyArgGlyProSerProGluGlySerSerSerThrGluSer 689

QY 1528 -----CCGCCCGCTTACGCGCGGATGTATAGAGGCTGGGGCGGCTCAGCGG 1481

Db 690 SerProGluHisProProLysSerArg----- 698

QY 1480 CCGTGCCTGGACACAGCGCGCGCATGCTGGCCGGTGAACGCTGCTTGCCAGACGG 1421

Db 699 -----ThraAlaArgArgGlySerArgSerPro----- 708

```

QY 1420 GCGCGCGCTGACGACGAGATGATCCGCGCGCTGCGAGGCTGTGCAAG 1361
DB 708 -----
QY 1360 TGGCGGATCGCCACCGTGGCGGCTGCAACACACAGTGGCGGAATGCCGCT 1301
DB 709 -----GluProLysThrLysSerArgThrProProValArgArgSerSerArgSer 725
QY 1300 GCAGGTGCGCGCGCGCGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1241
DB 726 SerProGluLeuThrArgLysAlaArgLysSerArgSerArgSerArgSerArgSer 745
QY 1240 CCGCGGCTCCAGCGGGAACAGCCAGTCCGCGCAT-----AGTCCCA 1193
DB 746 ProGluThrArgSerArgThrProProAlaArgArgSerProSerValSerSerPro 765
QY 1192 GCAGCGCGGGAAGTGGCCACAGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 1133
DB 766 GluProAlaGluLys-----SerArgSerArgArgArgSerArgSerArgSerPro 783
QY 1132 CGGAGCGGATGATGCTG-----GGCGAAGCGAGATGCGGCGGCGGCGGCTGT 1079
DB 784 ArgThrLysThrThrSerArgArgLysArgLysSerProLysProAlaGlyLeuGln 803
QY 1078 CCATCATAGATAGATGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1019
DB 804 ArgSerArgSerArgSerArgArg-----GluLysThrArgThr 816
QY 1018 TGGCGCGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 959
DB 817 ThrArgArgArgArgArgSerArgSerArgSerArgSerArgSerArgSerArg 835
QY 958 ACAGAGTATGATGATGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 899
DB 836 -----Ser 836
QY 898 CGCGCGGATGAGGATACGCGTCCAGCGAGCGAGCTGCGCGCGGCGGCGGCGGCG 839
DB 837 ArgSerArgSerArgValThrArgArgArgArgLysSerArgLysSerArgSer 856
QY 838 GCGGCA-----GGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 788
DB 857 ProAlaArgGlnLysSerArgThrSerArgArgArgArgArgLys----- 872
QY 787 GGTGCGAGATCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 728
DB 873 ArgSerArgThrProThrSerArgLysArgSerArg-----SerArgThrSerProAla 891
QY 727 CGCGCGAGATGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 677
DB 892 -----ProThrLysArgSerArgSerArgAla 900
QY 676 -----CATGCGCGCTACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 629
DB 901 SerProAlaThrHisArgArgSerArgSerArgThrProLeuLysSerArgArgSer 920
QY 628 CCGGCGGCGGCGG-----GGGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCG 578
DB 921 ArgSerArgThrSerProValSerArgArgArgSerArgSerArgThrValThrArg 940
QY 577 TGGCCATGCTCTCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 518
DB 941 -----ArgArgSerArgSer 945
QY 517 GGTCAAGTCCGCGCATGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 458
DB 946 ArgAlaSerProValSerArgArgArgSerArgSerArgThrProProValThrArg 965
QY 457 TGAAGCGCGCGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 422
DB 966 ArgSerArgSerArgThrProThrArgArgArgSerArgSerArgThrProProVal 985

```

---

```

QY 421 AGTCGCGGAGCGCTCGAAGCGGTAAGACCGGCTTCTGCG-----ACGAGTCCA 371
DB 966 ThrArgArg-ArgSerArgSerArgThrProProValThrArgArgArgSerArgSera 1005
QY 370 GGGGCGCGGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 324
DB 1005 rGThSerProLeuThrArgArgArgSerArgSerArgThrSerProValThrArgArg 1025
QY 323 -----CAATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 278
DB 1025 rSerArgSerArgThrSerProVal----- 1034
QY 277 TCATGTCGCGAGCGCTGCGAGGATGCTGCTGCGGCGGCGGCGGCGGCGGCG 218
DB 1035 -----ArgArgArgSerArgSerArgThrSerProVal-ThrArgArgArgSerArg 1051
QY 217 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 176
DB 1052 SerArgThrProProAlaLeuThrArgArgSerArgSerArgThrProLeuLeuProArg 1071
QY 175 GCGGCGTCC-----GACAGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 126
DB 1072 LysArgSerArgSerArgSerArgSerProLeuAlaLeuArgArgArgSerArgThrPro 1091
QY 125 AGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 66
DB 1092 ArgThrAlaArgGlyLysArgSerLeu-----ThrArgSerProProAlaLeuArg 1108
QY 65 AGG-----TCGAGAGCGGCGGAGATGCGGATGGAACATGCAATCTCTCTCGGCGATC 12
DB 1109 ArgArgSerArgAlaSerArgSerSerSerArgSerArgSerArgThrProProAlaThr 1128
QY 11 AAG 9
DB 1129 Arg 1129

```

---

```

RESULT 13
D70928
hypothetical protein Rv2913c - Mycobacterium tuberculosis (strain H37RV)
C:Date: 17-Jul-1998 #sequence_rev1501 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: D70928
R:Colo, S.T.; Broesch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gor,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete g.
A:Reference number: A70500; MIMD:98295987; PMID:9634230
A:Accession: D70928
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-611 <COL>
A:Cross-references: GB:174024; GH:AL123456; NID:93250700; PIDN:CAA98380.1; PID:el3
C:Genetics:
A:gene: Rv2913c

```

---

```

Alignment Scores:
  1,34e-08
Score: 294.00
Percent Similarity: 36.48
Best Local Similarity: 24.06
Query Match: 8.91%
DB: 2
Gaps: 25

```

---

```

US-10-009-782-1 (1-1758) x D70928 (1-611)
QY 58 TTGACATGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 117
DB 16 TyraSpValIleIleArgSpLysLeuThrPheArgSpLysThrGlyAlaAlaProLeuThr 35
QY 118 GCCGACCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 171

```

```

Db 36 ArgThrLeuGlyIleArgAspGlyValValAlaThrValAlaAlaGlyAlaIleuAspGlu 55
QY 172 GCGCCCGCGCACACCGCGGTGACGTGCGGGCGGTGCGCGCGCGGTCAATCGAC 231
Db 56 ThrGlyCysProGluValAlaValAlaAlaGlyLysTrpValValProGlyPheIleAsp 75
QY 232 TCGCACACCGCGCACACACTACTGCTGAGGGGTGCGCGACGACGACCAATCTCG 291
Db 76 ValHisThrHisTrpAlaPheAlaGluValIleuLeuAspProGlyLeuArgGluSerValArg 95
QY 292 CAGGCGCTCACACCGGTGCTACGCGCAATGCGCATGCGCGCGCGCGCGCGCGCAC 351
Db 96 HisGlyValThrThrValIleuLeuGlyAsnGlySerLeuSer-----ThrValTyr 112
QY 352 GCGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
Db 113 AlaHisSerLeuAspAlaAlaAlaAspLeuPheSerArgValGluAlaValProArgGluPhe 132
QY 391 -----GCTCTTACCGT-----TTGAGCGCGTTGCGCGCATCGCGTGGAGCGG 432
Db 133 ValIleuGlyAlaLeuArgAspAsnGlnThrTrpSerThrProAlaGluTyrIleGluAla 152
QY 433 TTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492
Db 153 IleAspAlaLeuProIleuGlyProAsnValSerSerLeuLeuGlyHisSerAspLeuArg 172
QY 493 GCGCGGTCTATGCGCGACTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534
Db 173 ThrAlaValLeu---GlyLeuAspArgAlaThrAspAspThrValArgProThrGluAla 191
QY 535 GAAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 594
Db 192 GluLeuAlaIleMetAlaIleLeuLeuAspGluAlaLeuGlyMetLeuGlyMet 211
QY 595 TCGACCGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 647
Db 212 Ser-----GlyMetAspAlaAlaIleAspLysLeuAspGlyAsp 224
QY 648 -----CGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 695
Db 225 ArgPheArgSerArgAlaLeuProSerThrPheAlaThrTrpArgGluArgGlyLeu 244
QY 696 CATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 755
Db 245 IleSerValLeuArg-----HisArgGlyArgIle-----LeuGlnSerAlaPro 259
QY 756 CGAGTGTGCGCGCGCGGTGTATCTCGCACACGATGATGCGCGCGCGCGCGCG 815
Db 260 -AspValAspAsnProValSerAlaLeuPhePheLeuAlaSerSerArgIlePheAs 279
QY 816 CCGCTCGCGCG-----GAGACGCGTGGCGCGT 839
Db 279 nArgArgLysGlyValArgMetSerMetLeuValSerAlaAspAlaLysSerMetProLe 299
QY 840 GATC-----GAGGC 848
Db 299 uAlaValHisValPheGlyLeuGlyThrArgValLeuAsnLysLeuLeuGlySerGlnVal 319
QY 849 CGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885
Db 319 LArgPheGlnHisLeuProValProPheGlnLeuTyrSerAspGlyIleAspLeuProVal 339
QY 886 -----CCCTACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 923
Db 339 LpheGlnGluPheGlyAlaGlyThrAlaAlaLeuHisLeuAspGluLeuGlnArgAs 359
QY 924 CGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945
Db 359 nGluLeuLeuAlaAspArgSerTyrArgSerPheArgArgGluPheAspArgIleTyr 379
QY 946 -----ATCATCACCCTGCTGCAAGCC 965

```

```

Db 379 sLeuGlyProSerLeuThrHisArgAspPheHisAspAlaValIleValGluCys---Pr 398
QY 966 CTTCCCGCACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1025
Db 398 oAspLysSerLeuIleGlyLysSerPheGlyAlaIleAlaAspGluArgGly----- 415
QY 1026 GTACAGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1078
Db 416 -----LeuHisProLeuAspAlaPheLeuAspValLeuValAspAs 429
QY 1079 ACAGAACCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1138
Db 429 nGlyGluArgAsnValArgTrpThrThrIleValAlaAsnHis-----ArgPr 445
QY 1139 TGCCGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1195
Db 445 oAsnGlnLeuAsnLysLeuAlaAlaGluProSerValHisMetGlyPheSerAspAlaG 465
QY 1196 GGCATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1224
Db 465 yAla-HisLeuArgAsnMetAlaPheTyrAsnPheGlyLeuArgLeuLysArgAla 485
QY 1225 -----CCGCTGAGACCGCGGTATGAGATGA 1252
Db 485 rAspAlaAspArgAlaGlyGlnProPheLeuSerIleGluArgAlaValTyrArgLeu 505
QY 1253 CCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1312
Db 505 nGlyGluLeuAlaGluTrpPheGlyIle---GlyAlaGlyThrLeuAspGlnGlyAsp 524
QY 1313 TCGCGCACTGTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1372
Db 524 rGluAspPheAlaValIleAspProThrHisLeuAspGluSerValAspGlyTyrHisG 544
QY 1373 CT-----ACCGACCGCGCGCGCGCGCG 1393
Db 544 LuGluAlaValProTyrTyrGlyGlyLeuArgArgHeValAsnArgAsnAspAlaThrV 564
QY 1394 TCCATTCGCGTACGTCACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1452
Db 564 aValAlaThrGlyValGlyGlyThrValValPheArgGlyGlnPheGlyGlnP 584
QY 1453 -----CATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1486
Db 584 hArgAspGlyTyrGlyGlnAsnValLysSerGlyArgTyr----- 597
QY 1487 GAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1530
Db 598 -----LeuArgAlaGlyGluLeuGlyAlaAla 606

```

RESULT 14  
 A:3932  
 mucin 2 precursor, intestinal - human (fragments)  
 N:Alternate names: mucin SMUC-41  
 C:Species: Homo sapiens (man)  
 C:Date: 10-Mar-1993 #sequence\_revision 12-Apr-1996 #text\_change 05-Nov-1999  
 R:Accession: A49963; A45106; B45106; B3532; B61257; P00328; P00329  
 R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
 J. Biol. Chem. 269, 2440-2446, 1994  
 A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of  
 A:Reference number: A49963; MUID:94132002; PMID:8300571  
 A:Accession: A49963  
 A:Molecule type: mRNA  
 A:Residues: 1-639 <GDI>  
 A:Cross-references: GB:L21998  
 R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.  
 J. Biol. Chem. 267, 21375-21383, 1992  
 A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up  
 A:Reference number: A45106; MUID:93016075; PMID:1400449  
 A:Accession: A45106  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 626-1895 <GDI>

```

Db 1403 ThrThrThrProSerProProThrThrThrThrLeuPro-----Pro 1411B
QY 132 ACTGCGGCGTCCGGCGGACCGCATCGCGCATCGGCATCTGTGGACGCCGCCGC 181
Db 1419 Thr-----ThrThrProSerProProThrThrThrThrProProPro 1430
QY 182 ACACCCGGGTGCAGCTGCGGGCTGGTGGTCGGCCGCTTCATCATCATCGACACC 241
Db 1435 ThrThrThrPro-----SerProThrThrThrThrThrPro 1448
QY 242 ACGAGCAACTACTGCTCAGCGCGTGGGAGATGACGCCAAGATCTCGAGGCGCTCA 301
Db 1449 LeuProThrThr-----ProSer 1455
QY 302 CCAAGGGGTGCAGGGCAATTGGGCATACGCTGGCGCGGTGGCGACGCCACCGC 361
Db 1456 ProProIleSer-----ThrThrThrThr 1463
QY 362 CCGCCCCCGTGACCTGCTGAGCAAGCGGCTCTTACCGTTTGAGCGCTTCGCCGACT 421
Db 1464 ProProProThr-----ThrProSerProPro 1473
QY 422 ACCTGAGCGCTTGGCGGCACGCGCGCGCTGACAGCGCGCTGATATGTGGCCATT 481
Db 1474 ThrThrThrProSerProProThrThrThrProSerProProThr----- 1489
QY 482 CAACGTCGCGCGCGCGTCTATGGCGGACTTGCAGCGCGCGCGCACGAGAGAAATCG 541
Db 1490 -----ThrThrThrThrProProProThrThrPro 1500
QY 542 CGGCATCGGGACCTGGCGCGAGACCATGGCGAGCGCGCATTTGGACCG 601
Db 1501 SerProProweThrThrProIleThrPro-----Pro 1511
QY 602 GCGCCTTCTACCGCGCGCGCGCGCGCCACCAACGAGATCATCATGATGTGCGCGC 661
Db 1512 AlaSerThrThrThrLeuProProThrThrPro-----Pro 1523
QY 662 CGCTGACGGCGATGGCGGCATGCCACCAACCATTTGGCGACGAGCGACGACAATCG 721
Db 1524 -----SerProProThrThrThrThr-----ThrThr 1532
QY 722 TGGCGCGCGTGGAGAAACCTTCCGCATCGCGGAGCTGGATCCCT--AGTGGCG 898
Db 1533 ProProProThrThrThrProSerProProThr-----ThrThr 781
QY 782 CGCAGCACAAAGTATGGCGCGACGCCAATTTCGGCGCTCGCGGAGACGTCGCCGTGA 841
Db 1544 -----ThrThrThr----- 1548
QY 842 TCGAGCGCCCGCATGCGCGCGCCAGAGCATTCGTGAGCGGTATCCCT--AGTGGCG 898
Db 1549 -----ProPro-----ThrSerThrThrLeuProProThrThrPro 1562
QY 899 GCTCGACCATGCTCAGCAGAGACCGCGTGGTGGCGGACGACCATCATCACTGTGT 958
Db 1563 SerProProProThrThrThrThr-----ProProProThr 1575
QY 959 GGAACCGCTTCCCGAAGTACGGCGGCGACCTGTGATGAGTGGCGGAGCGCGGA 1018
Db 1576 ThrThrThrPro-----ProThrThrThr 1584
QY 1019 AATCAAGT-----ACGAGCTGTGCCCGAGCTGCACCGCGCGCGCGCA 1063
Db 1585 ThrProProProThrThrThrThrThrThrProProProThrThrThrProSerPro 1604
QY 1064 TCTACTTCATGATGAGCAAC-----CCGAGTCGCGGAGATTCGTGGCGTTCGCCCGA 1117
Db 1605 ProThrThrThrThrThrProProProProThrThrThrProSerProProThrThr 1624
QY 1118 CCAATGATGGCTTCAGAGCGCGTGGCGGACGAGAGACGCCGATCCGCGCTGTGGGCA 1177
Db 1625 ProIleThrProProThrSer-----ThrThr----- 1632

```



```
QY 784 GCGAGATACCA---CCGGCAGCTCCAGCTCGCGCCGATCGGAGATTTCCTCCAGCG 728
Db 731 ArgLeuGlyProGlyProAlaArgGlyProHisGlyArgAspGlyArg----- 746
QY 727 CCGCCACGATGTCGCTCGCTTCGTCGCGCATGTGGGTGGCTAGA----- 683
Db 747 -----ValArgLeuArgGlnValHisAlaAspGlnArgAspProLeuGlnGly 762
QY 682 -----TGGCGCATGGCGCTCCAGCGCGCGCCGACACCTCGATGATCT 641
Db 763 ProGlyGlnProAlaGlnArgArgGlnAlaArgAla-ArgProAlaProVal----- 779
QY 640 CTCGCTGGGTGGCGCGCGCGCGGTAGAGCGCCGCT-----CGAATGCCGA 587
Db 780 -----ArgGlyGlyThrArgAlaProGlyGlnGlyProAlaArgGlyProGly 795
QY 586 TGGCGCCGCTGGCCATGGCTCTCTC-----GGCCAGGTC----- 553
Db 795 uProHisArgAlaHisAlaThrLeuGlnProArgHisLeuHisGlyArgValArgArgAs 815
QY 552 -----CCGATGGCGCGCATTTCTCTCGCGGTGGCGCGCGCT 515
Db 815 PProGlnAlaLeuArgGlyAspProArgGlyGlnGlyProGlyLeuProAlaGlyProVa 835
QY 514 GCAAGTCCGCGCATGACCGCGCGCGCGCATGGAATGGCCACCATACAGCGCGCGTTGA 455
Db 835 lLeuLeuGlnHisGlnGlyArgAlaLeuArg-----GlyValArg 848
QY 454 CCGCGCGCGCGGTGGCGCGCGCATGAGTGGCGGACCGCTCGAAGCGTAAAG 395
Db 848 gGlyArgArgHisAlaGlnAspArgArgAspGlnProAlaGlyArgLeuArgAlaValArg 868
QY 394 AGCGGCTCTCTCCAGCAGCTCCAGGGGCGC-----GGCGCGGTGGCGTGGCGCA 344
Db 868 gGlyValProArgGlyProValGlnProGlyAspAlaArgGlyHisLeuGlnGlyGlnGly 888
QY 343 GCGCGCGCGCATGATGCGCGCATGTCGCGTACCGCGTGTGACGCGCTCGAGATCT 284
Db 888 uHisArgArgGlyProArgHisAlaAspArgGlyGlyArgGly---LeuLeuGlnArgGly 907
QY 283 TGGCGCTCATGTCGCGCAGCGCTGAGCAGGTAGTGTCTGCTGGGTGGCGATGCA 224
Db 907 lHisProHisLeuAlaValPro--GlyHisAlaArgArgArg----- 921
QY 223 AGCGGCGCGCGACCGCGCGCGCATGTCGACCGCGTGTGCGCGCGCGCGTCCGACA 164
Db 922 SerGlyLeuArgProSerGlyProAlaArgHisHisAlaLeuGlyArgArgGlyProAla 941
QY 163 GATCGCGATGGCGCGCATGCGTCCGCGCA-----CGCCCA----- 125
Db 942 -----ArgGlnAlaGlyGlyArgAlaProGlnAlaLeuGlnArgProHisHis 957
QY 124 -----GATCGCGCGCGCGCGCGGTGTCGCGCGCATGAGGTCGCGCGCGCA 71
Db 958 LeuArgAlaGlyArgAlaAspHisGlyValAlaLeuArgArg-----HisProGln 974
QY 70 GCAAGCAGGTGGAAGGCTGGGAATCGGATGGGACATGGAATCTCTCCGCGATCA 11
Db 975 AlaProAlaArgAlaSer-----ValProAlaArgGly 984
QY 10 AGTGA 5
Db 985 GlnGly 986
```

Search completed: May 11, 2003, 12:12:51  
Job time : 84.5 secs



GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_nzp model

Run on: May 11, 2003, 12:12:55 ; Search time 77 Seconds

(without alignments)  
7808,619 Million cell updates/sec

Title: US-10-009-782-1

Sequence: 1 gaattcactgacgcgga.....ccctgagctacgagaagctt 1758

#### Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 787003 seqs, 171007862 residues

Total number of hits satisfying chosen parameters: 1574006

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+npz.model -DEV-xlp  
-O/cgnt2.1/USPTO.spool/US10009782/runat\_07052003\_122519\_23229/app\_query.fasta.1.1927  
-DB-Pending Patents\_AA.New -QWMT-Itasca -SUFFIX-rapn -MINMATCH-0.1 -LOOPEXT-0  
-LOOPEXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-human40.cdl  
-LIST-45 -DOCALLIGN-200 -THR\_SCORE-pct -THR\_MAX-100 -THR\_MIN-0 -ALIGN-15  
-MODE-LOCAL -OUTWMT-pto -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000  
-USER-US10009782.ecgn.1.1.42.gunat.07052003\_122519\_23229 -MCPU-6 -ICPU-3  
-NO\_XLPEXT -NO\_MMAP -LARGEQUERY -NEG\_SCORES-0 -WAIT -LONGLOG -DEV-TIMEOUT-120  
-NARN\_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -Fgapop-6 -Fgapext-7  
-YGAPOP-10 -YGAPEXT-0.5 -DELop-6 -DELext-7

#### Database :

Pending\_Patents\_AA.New:\*  
1: /cgnt2.6/ptodata/2/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgnt2.6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgnt2.6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgnt2.6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgnt2.6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgnt2.6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgnt2.6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1111.5	33.7	475	6	US-10-282-122A-50211
2	1106.5	33.5	479	6	US-10-282-122A-60214
3	659	20.0	530	6	US-10-156-761-10819
4	539	16.3	576	6	US-10-156-761-11333
5	455	13.8	2294	6	US-10-366-683-17231
6	455	13.8	2294	6	US-10-366-683-17231
7	434	13.2	885	6	US-10-419-128-17231
8	434	13.2	885	6	US-10-366-683-26129
9	433	13.1	937	6	US-10-419-128-26129
10	433	13.1	937	6	US-10-366-683-19446
11	432.5	13.1	663	6	US-10-366-683-30843

12	432.5	13.1	663	6	US-10-419-128-30843	Sequence 30843, A
13	431.5	13.1	822	6	US-10-366-683-21920	Sequence 21920, A
14	431.5	13.1	822	6	US-10-419-128-21920	Sequence 21920, A
15	430.5	13.0	651	6	US-10-366-683-32204	Sequence 32204, A
16	430.5	13.0	651	6	US-10-419-128-32204	Sequence 32204, A
17	424.5	12.9	1002	6	US-10-366-683-27980	Sequence 27980, A
18	424.5	12.9	1149	6	US-10-419-128-27980	Sequence 27980, A
19	424.5	12.9	1149	6	US-10-366-683-25557	Sequence 25557, A
20	424.5	12.9	1149	6	US-10-419-128-25557	Sequence 25557, A
21	423	12.8	801	6	US-10-366-683-29274	Sequence 29274, A
22	423	12.8	801	6	US-10-419-128-29274	Sequence 29274, A
23	421.5	12.8	863	6	US-10-366-683-26099	Sequence 26099, A
24	421.5	12.8	863	6	US-10-419-128-26099	Sequence 26099, A
25	413.5	12.5	774	6	US-10-366-683-16789	Sequence 16789, A
26	413.5	12.5	774	6	US-10-419-128-16789	Sequence 16789, A
27	410	12.4	1706	6	US-10-366-683-31760	Sequence 31760, A
28	410	12.4	1706	6	US-10-419-128-31760	Sequence 31760, A
29	408.5	12.4	638	6	US-10-366-683-27068	Sequence 27068, A
30	408.5	12.4	638	6	US-10-419-128-27068	Sequence 27068, A
31	407.5	12.4	672	6	US-10-366-683-16941	Sequence 16941, A
32	407.5	12.4	672	6	US-10-419-128-16941	Sequence 16941, A
33	407	12.3	639	6	US-10-366-683-20571	Sequence 20571, A
34	407	12.3	639	6	US-10-419-128-20571	Sequence 20571, A
35	404	12.2	1053	6	US-10-366-683-26140	Sequence 26140, A
36	404	12.2	1053	6	US-10-419-128-26140	Sequence 26140, A
37	402.5	12.2	1418	6	US-10-366-683-32367	Sequence 32367, A
38	402.5	12.2	1418	6	US-10-419-128-32367	Sequence 32367, A
39	401.5	12.2	639	6	US-10-366-683-28453	Sequence 28453, A
40	401.5	12.2	639	6	US-10-419-128-28453	Sequence 28453, A
41	399	12.1	724	6	US-10-366-683-31715	Sequence 31715, A
42	399	12.1	724	6	US-10-419-128-31715	Sequence 31715, A
43	395	12.0	802	6	US-10-366-683-25050	Sequence 25050, A
44	395	12.0	802	6	US-10-419-128-25050	Sequence 25050, A
45	394	11.9	631	6	US-10-366-683-30909	Sequence 30909, A

#### ALIGNMENTS

RESULT 1  
US-10-282-122A-50211  
Sequence 50211, Application US/10282122A  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELTRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
PRIOR FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931

```

Db      231  |||SerHisLeuLeuIscyAlaIaG|YProSerHisnTrpG|YAArgSerAlaIaG|ValLeuAla 250
OY      838  CTGATCGAGGCCCGCCATGCGCGCGCGGACGAGCATCTCGCTGGAGCGGGTATCCCTACGTGGCC 897
Db      251  SerLeuGluAArgAlaArgArgIleG|LnProValIeIlyAspCyStrYTrProTYSerArg 270
OY      898  GGCCTCACCATCGATCGCAAGACAGACGCCGCTGCTGCTGGCGCGGACGACACATCATCACCTGG 957
Db      271  SerSerSerTrpLleAspLleIleG|LysAlaIaThrG|YAspLleAspLleIleThrTrp 290
OY      958  TGCAGAGCCCTTCCCGGACTGAGCGGGCGGAGCCTGGATGCAATCCCGCGCGACCGCGGC 101
Db      291  SerAspProHisProGluMetAlaG|LysLeuLeuLdnArgThrLleAlaIaG|LdnTrpArg 310
OY      1018  AAATCCAGATACGACGGTGGTGGCCGAGCTGACAGCCCGCGCGCGCCATCTCACTCATGATG 1077
Db      311  ValSerGluGlnAspAlaAlaAlaArgArgLeuGlnProAlaG|YAlaValYTrHisnMet 330
OY      1078  GACGACCCCGCATGCGAGCGGCATCCGAGGGCTTCGCGCGACCATGATCGGCTCGACGCGC 1137
Db      331  SerGluAspAspAlaArgArgIleLeuSerHisProAlaIaThrMetValG|YSerAspLys 350
OY      1138  CTGCGGCACGACGAGCGCGCCGCGCATCCGCGCTGTGGGCGACCTTCGCGCGGGTCTGTGGGG 1197
Db      351  LeuProHisnAspProLdnProHisProArgLeuTrpG|YAlaIaPheProAlaValLeuG|Y 370
OY      1198  CACTATGCGCGCGACCTGGGCGCTGTCCCGCTGCGAGACGGCGGTATGGAGATGACCGCGC 1257
Db      371  TyrTrpAlaArgAspLdnArgLeuIleSerLeuGlnAspAlaValAlaArgLysMetThrAla 390
OY      1258  CTGACCGCCCGCGCGCTGGCGCTGGCGCGCGCGCGGACGGCGAGCGCGCGGCTACTTCGCC 1317
Db      391  LeuSerAlaAlaArgArgPheG|YLeuAlaAlaArgArgIleValAlaArgValG|YTrHisAla 410
OY      1318  GACCTGGTGTGTTCGACCCCGCGCCACGCGTGCGCGCATCCGCCACTTCGCAACACCGCTAC 1377
Db      411  AspLeuValLeuPheAspAlaAlaAlaArgValArgAspAlaIaThrPheGlnGlnProGln 430
OY      1378  GAGCGCGCGCGCGCATCATCTCCGTGTACGTACGTACAGCGCGCGCGGTCTGGCAAGACAG 1437
Db      431  GlnProAlaHisG|YIleGlnAlaValIleTrpValAsnG|YAlaLeuSerTrpArgAspLys 450
OY      1438  GCGGTCAACGCGCCACGATGCGCGCGCGCGCTGCTGCGACGACGCGCGCGCGACCGCGCGC 1497
Db      451  AlaProThrG|YAlaArgAlaG|YAlaArgPheValAlaArg-----G|YLeuArg 466
OY      1498  CAGCCCTTCAATCCGCGC 1515
Db      467  AlaProAlaThrProG|Y 472

RESULT 2
US-10-282-122A-60254
/ Sequence 60254, Application US/10282122A
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liansu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELIPIA.034A
/ CURRENT APPLICATION NUMBER: US/10/282.122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21

```

```

PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 60254
LENGTH: 479
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-10-282-122A-60254

```

```

Alignment Scores:
Pred. No.: 2,63e-31 Length: 479
Score: 1106.50 Matches: 231
Percent Similarity: 63.29% Conservative: 69
Best Local Similarity: 48.73% Mismatches: 171
Query Match: 33.54% Indels: 3
Caps: 3

```

US-10-009-782-1 (1-1758) x US-10-282-122A-60254 (1-479)

```

61 GACCTGCTGTCGGCGGCGGACCCCTATGACGCGCAAGAACCCGCGGCGCGCGC 120
DB 4 AsptrleuhyelsasnValThrValIleAspGlySerGlyProGlnTrpArgAla 23
QY 121 GACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 24 AspValAlaValIleAspGlySerGlySerGlySerGlySerGlySerGlySer 43
QY 181 CACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 44 GluGlnValIleAspGlySerGlySerGlySerGlySerGlySerGlySerGly 63
QY 241 CACGAGCAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
DB 64 HisAspAspIleAsnValIleAsnValIleAsnValIleAsnValIleAsnVal 83
QY 301 ACCACGGTGTGACGCGGCAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 357
DB 84 ThrThrValIleValIleValIleValIleValIleValIleValIleValIle 103
QY 358 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
DB 104 ValProAspProMetAspIleuSerGlyGlnGlnGlnHisPheIleThrPro 123
QY 418 GACTACCTGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 477
DB 124 AlaTrpAlaHisAlaValAlaValAlaValAlaValAlaValAlaValAlaVal 143
QY 478 CATTACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 537
DB 144 HisThrAlaIleuArgAsnHisMetAspIleuSerGlyProAlaAsnGlnThr 163
QY 538 ATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597
DB 164 IleAlaGlyMetArgValAlaIleuArgAspIleuArgGlnGlnAlaIleuSer 183

```

```

QY 598 ACCGCGGCTTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657
DB 184 ThrGlyLeuAlaValAlaSerAlaPheGlnSerThrThrGlnValIleAlaIle 203
QY 658 CGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 717
DB 204 GluGlnLeuAlaIleGlySerGlyValIleThrThrHisMetAspSerGlyPro 223
QY 718 ATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 777
DB 224 IleLeuGlnAlaIleuAspIleuSerArgIleGlyArgHisGlnValProVal 243
QY 778 ATCTGCGCACCAAGGTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 837
DB 244 ValSerHisHisIleValAlaValAlaValAlaValAlaValAlaValAla 857
QY 838 CTGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 897
DB 264 PhePheAspGluMetArgGlnGlnGlnAlaSerIleAlaCysAspCysTrp 283
QY 898 GGTTCACACATGCTCAAGCAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 957
DB 284 SerSerSerThrIleuAspMetIleValIleValIleValIleValIleVal 303
QY 958 TGCAAGCCCTCCCGCACTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1017
DB 304 SerGlnAlaGlnProGlnGlnAlaGlyIleThrIleuGlnIleAlaAspGln 323
QY 1018 AAATCCAAAGTACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1077
DB 324 ValSerIleuHisAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 343
QY 1078 GACGACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1137
DB 344 AspGlnGlnAspValArgArgValIleMetArgTrpProValIleMetIle 363
QY 1138 CTGCGCGACGACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1197
DB 364 LeuProAspAspProMetProHisProAlaGlnProAlaPheProAlaIle 383
QY 1198 CACTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1257
DB 384 HisTrpSerArgAspIleuGlnIleuPheProIleuThrThrAlaValHis 403
QY 1258 CTGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1317
DB 404 LeuSerAlaAlaArgPheGlnIleuAlaAspArgGlyLeuValIleGlyTr 423
QY 1318 GACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1377
DB 424 AspLeuValIleuPheAspProGlnThrValArgAspValAlaSerPheSer 443
QY 1378 GAGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1434
DB 444 ArgProAlaAspGlyIleGlnAlaValIleValAlaValIleValAlaVal 463
QY 1435 CAGCGCTTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1476
DB 464 LysLysIleThrGlnArg--AlaGlyArgPheLeuArgArg 476

```

```

RESULT 3
US-10-156-761-10819
Sequence 10819, Application US/10156761
GENERAL INFORMATION:
APPLICANT: OMOIRA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN.
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262

```

CURRENT APPLICATION NUMBER: US/10/156,761  
 CURRENT FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 10819  
 LENGTH: 530  
 TYPE: PRT  
 ORGANISM: Streptomyces avermitilis  
 US-10-156-761-10819

## Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:
9e-16	659.00	530	182	71	201	92
Percent Similarity:	46.34%					
Best Local Similarity:	33.33%					
Query Match:	19.98%					
DB:	6	gaps:	12			

US-10-009-782-1 (1-1758) x US-10-156-761-10819 (1-530)

QY 61 GACCTGCTGCTGCGGGGCGACCTCATGCGGCGACCAACCCGCGGGCGCGCC 120  
 DB 2 AspleuValIleArgAspValArgValAlaAspGlyThrGlyAlaSerCysArgAla 21  
 QY 121 GACCTGCGGCGGGGCGGCGGCGCATGCCCGCATCGGCGATGTCGTGGAGCGCGCGC 180  
 DB 22 AspValGlyIleThrGlyAlaGlyAlaGlyIleAlaGlyIleArgAspGlyProAlaLeuGly 41  
 QY 181 CACACCCGGGTC--GACGTGTCGGGCTGTGGTGGCGGCGGCGCTTCATGATCGCAC 237  
 DB 42 GlyArgValIleuAspAlaGlyGlyIleuAlaLeuAlaProGlyPheIleuAspMetHis 61  
 QY 238 ACCGAGCGACACTACTCTCAGCGCTCGGCGATGACGACGCGGCGGCGGCGGCGGCG 297  
 DB 62 AlaHisSerAspLeuAlaLeuAlaArgAspProAspHisSerAlaLysAlaIleGly 81  
 QY 298 GTCAACGAGGTGTACGCGCATTTGGCGATCGGCGGCGGCGGCGGCGGCGGCGGCG 345  
 DB 82 ValThrLeuGluValValGlyGlyIleuSerGlyLysProValAspAspArgThr 101  
 QY 346 -----GCGCAGCCCAACCGCGCGCGCC 369  
 DB 102 LeuAlaGluValArgAlaAlaIleThrGlyTyrPasnGlyIleGly----- 116  
 QY 370 CTGAGCTGCTGCGGCGGCGGCTGTACGCTTTCAGCGCTTCGCGACTGAC 429  
 DB 117 ---AspAspIleAsp-----PheThrTyrArgSerValGlyGlyTyrIleuAsp 131  
 QY 430 GCGTTGGGCGGCGGCGGCGGCGCTCAAGCGCGCTGTATGGTGGCGCATTCACGCTG 489  
 DB 132 ArgLeu---AspGlnGlyIleAlaValAlaIleAlaIleTyrIleuLeuProGlnGlyThrVal 150  
 QY 490 CGGCGCGGCGTATGCCGAGACTTGACGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCG 549  
 DB 151 ArgMetLeuAlaValAlaIleTyrPasnArgAlaAlaIleThrProGlnGlyIleuAspArgMet 170  
 QY 550 CGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 609  
 DB 171 ArgArgLeuValAlaAspIleMetArgGlyGlyAlaValGlyMetSerSerGlyIleuThr 190  
 QY 610 TACCG 669  
 DB 191 TyrThrProGlnMetTyrAlaGluAspAlaGluLeuThrGlnGlyIleuSerArgValAla 210  
 QY 670 GCGCATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 729  
 DB 211 GluTyrGlyGlyTyrTyrCysProHisIleArgSerTyrGlyAlaGlyAlaIleGlnAla 230  
 QY 730 CTGAGGAAACCTCCGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 789

DB 231 TyrGluMetValAlaLeuThrArgGluAlaArgCysProLeuHisLeuAlaHisAla 250  
 QY 790 AAGTCATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 849  
 DB 251 ThrMetAspPheGlyValAlaAspGlyArgAlaProAspLeuAlaLeuAspGly 270  
 QY 850 GCGATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 906  
 DB 271 AlaLeuAlaGluGlyAlaAspIleThrLeuAspThrGlyProGlyThrProGlyCysThr 290  
 QY 907 ATGCTCAAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 963  
 DB 291 ThrIleuVal-----AlaMetLeuProSerTyrAlaGlyGlu 302  
 QY 964 CCCTCCCGGAA--CTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
 DB 303 GlyIleProGluAlaValAlaLeuAlaArgLeuAlaAspArgThrValAlaGlyIleArg 322  
 QY 1021 TCGAAGTACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1038  
 DB 323 HisAlaMetGluValValGlyAlaAspGlyCysAlaGlyValProIleGluTyrAspThr 342  
 QY 1039 -----CCGAGCTGCGAGCGG----- 1053  
 DB 343 IleGluIleSerGlyValAlaSerAspProAlaLeuAlaProGlyValGlySerThrValGln 362  
 QY 1054 -----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
 DB 363 AlaSerAlaAspLeuArgGlyGlyAlaProThrThrAlaArgArgLeuLeuAsp 382  
 QY 1081 -----GAAACCGAGCTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1101  
 DB 383 AspArgLeuGlyCysThrIleLeuGlnHisValGlyHisGlyIleGlnHisValArgAlaIle 402  
 QY 1102 CTGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1161  
 DB 403 MetArgHisArgValHisThrGlyGlySerAspGlyIleLeuGlnGlyThrIleAspProHis 422  
 QY 1162 CGGCGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1221  
 DB 423 ProAlaGlyAlaTyrGlyIleThrPheProHisTyrLeuGlnHisTyrValArgIleuGlyAla 442  
 QY 1222 TTCGCGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1281  
 DB 443 LeuSerLeuGluGluCysValAlaHisLeuThrGlyArgProAlaAlaArgLeuArgLeu 462  
 QY 1282 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1341  
 DB 463 ProAspArgGlyLeuValArgGluGlyTyrArgAlaAspLeuValLeuPheAspProSer 482  
 QY 1342 ACGTGCGCGATACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1401  
 DB 483 ThrValAlaAlaGlySerThrPheGlnAspProArgArgLeuProAlaGlyLeuProHis 502  
 QY 1402 GTGTAAGTCGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1461  
 DB 503 ValLeuIleAspGlyArgPheValAlaGluAspGlyArgArgThrAspValLeuAlaGly 522  
 QY 1462 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1479  
 DB 523 ArgAlaValArgArgThr 528

RESULT 4.  
 US-10-156-761-11333  
 Sequence 11333, Application US/10156761  
 GENERAL INFORMATION:

APPLICANT: OMIURA, SATOSHI  
 APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIKAWA, JUN  
 APPLICANT: HORIKAWA, HIROSHI  
 APPLICANT: SHIBA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHIYUKI  
 APPLICANT: HATTORI, MASAHIRA

QY 697 ATGCGCGAGAGCGAGACACATCTGGCGCGCGCTGGAGAAACCTTCGCATCGCGCGC 756  
 Db 233 -----ThgInIleGuaIleValaIleAlaIleAgIleCysIleuAspIlePhe-----Ser 247  
 QY 757 GAGCTGGACGCGCGCGTGGTGAATCTGGACACCAACAGCTATGGCGACCC--AAATTG 813  
 Db 248 AspAlaGluIleAspIleuPheValGluMetSerAlaAlaAlaGlyAlaArgProIleuAsnTrp 267  
 QY 814 GGC----- 816  
 Db 268 AsnValIleuThrIleAspAlaAlaValProGluArgValProIleuArgIleuThrIleAsn 287  
 QY 817 ---GCGTGGCGCGAG-----ACGTCGCGCTGAATGAGCGCC 849  
 Db 288 GluArgAlaArgIleGlyAlaGlyIleArgValValAlaIleuThrMetProIle----- 304  
 QY 850 GCCATGGCGCGCGAGACGCTGCTGCGTGAAGCGTAT----- 885  
 Db 305 ---IleuThrProMetAsnMetSerIleuGlyThrPheCysAlaIleuAsnIleuLeuProGly 323  
 QY 886 -----CCCTTAC 891  
 Db 324 TrpIleGluValIleuGlyIleuProValProGluArgIleAlaIleArgIleuArgAspProAsn 343  
 QY 892 GTGGCGCGCGCTCCACCATGCTCAAGAGAGC-----CGCGTG 927  
 Db 344 ValAlaArgAlaGluIleuIleuArgIleArgAlaAspSerIleGluAlaGlyAlaPheArgAlaGlu 363  
 QY 928 CTGCTGGCGCGAGACCACTCATC-----ACGTGGTCAAGCGCTTCCCGCACTGAGC 981  
 Db 364 AlaAsnPheIleArgIleArgIleIleAlaIleAspThrIleThrAlaAsnGluIleuThr 383  
 QY 982 GGGCGGCACTGATGAATGATCGCGCGCGAGCGGCGGAAATCCAAAGTACAGCTGGTCCC 1041  
 Db 384 GlyArgValValaAsnAspIleAlaIleAlaGluIleuGlyIleAsnProPheAsnCysValIleuVal 403  
 QY 1042 GAGCTGCACCGCGCGCGCCCATCTACTCATGATG-----GACGAA 1083  
 Db 404 GluIleCysAlaIleAspIleuArgIleThrValIleuThrProMetProThrAspAsnAsp 423  
 QY 1084 CCGGAGTGCACGCGCTCTGGG-----TTGCGCCG 1116  
 Db 424 ProAspSerTrpAlaIleuArgAlaIleGluThrTrpArgHisIleAspValIleuIleuGly 442  
 QY 1117 ACCATGATCGCTCGACGCGCTGCGCGACAGACGCGCCGATCGCGCTGGGGGC 1176  
 Db 443 -----GlySerAspAlaGluAlaHisIleuAspArgIleMetCysGlyAlaIleProIleuThr 459  
 QY 1177 ACCTTCGCGCGGTCGGGCGCACTAAGCGCGCGACCTGGCTCTTCCGCTGAGAGC 1236  
 Db 460 Thr-----ArgPheIleuGlyAspCysIleuAlaGlyAlaGlyIleuValProIleuGluIle 477  
 QY 1237 GCGGATGAAGATGACGCGCTGACCGCGCGCGCGCTGCGCGCTGGCGCGCGCGGCGAG 1296  
 Db 478 AlaValIleuMetIleuThrAspAspProAlaIleArgIleuPheGlyIleuAlaGlyAspArgGlyArg 497  
 QY 1297 CTGGAGCGCGGCTACTCTCGCGACCTGGTGGTTCGACCCGCGCAGGTG--GCCGAT 1353  
 Db 498 IleArgIleuGluPheHisIleAspIleuValIleuPheAspProAlaArgIleAspAlaGly 517  
 QY 1354 ACCGCACTTCGCAACACCT-----ACGAGCGCGCGCC 1389  
 Db 518 LysAlaIleuThrIleuValHisAspIleuProGlyAspSerProArgIleuAspSerIleuAlaIle 537  
 QY 1390 GGCATTCATTCGCTGATGATCAACGCGCGCGCGCTGCGCAAGCAGCGGTACCGCGC 1449  
 Db 538 GlyAlaThrAlaValAlaTrpValAsnGlyAlaGluAlaIleArgAspAspValValThrGly 557  
 QY 1450 CAGCATGCGCGCGCGCTGCTC 1470  
 Db 558 AlaValProGlyIleuValIleu 564





```

Query Match: 13.798      Index: 281
DB: 6      Gaps: 46

US-10-009-782-1 (1-1758) x US-10-419-128-17231 (1-2294)

QY 9 CTTGATGCGGGAAGAGAAATTTCCATGTCCCAATCCGATTCGACCCCTTGACCTGCT 68
DB 675 LeuAlaIaThpProGlnAaPaaAlaIaThpProPheAlaIle-----ProAla 69
QY 69 GCTCGCGGCGGCGACCTCTCATCGACGGCAGCAACACCCGCGGCGCGCGCGACCT--- 122
DB 691 AlaValGlyIaThpGlyAaThpGlnAaThpProGlyLeuAlaAlaGlyLeuAlaValaThpAla 711
QY 126 -----GGCGCTGCGCGCGCA-----CCGATCGCGCGCGCATCGCGAGATCT 164
DB 711 AspGlyAlaGlyAaThpAaThpAaThpAlaGlyAlaGlnAlaIaThpProGlnAaThpAaThp 730
QY 165 GTCGCGAGC-----173
DB 731 ProGlyAlaGlyAlaThpProPaaProGlyAlaAlaAlaAlaGlyAaThpAlaIaThpPro 750
QY 174 -----CGCGCGCACACCCCGGCTGACGTCTCGGCGCGCGGCGGCGCGCGCG 221
DB 751 GlyAlaLeuProAaThpAaThpGlnAaThpProAlaGlyGlyGlyAlaThpAaThpAaThp 770
QY 222 CTTGATGCACTGCGACACCCAGACAGACACTCTGCTCAAGCGCGTGCAGATGACGC 281
DB 771 ProLeuGlyLeuAaPaaIaThpAaThpAaThpAlaValaGlnAaThpGlnAaThpAla 790
QY 282 CAAGATCTCGCAGGCGCTGACACACAGTGGTGCACGGGCAATTGCGGCAATCGACCTGCGCGC 341
DB 791 -----AaThpAaThpProAlaIa 796
QY 342 GCTGCGCGCA-----CGCCACCCCGCGCGCGCGCGCTGACCT---GCTGAGCAAGCGG 392
DB 797 AlaAlaAlaIaThpAaThpAaThpProAlaGlnGlyGlnGlyProGlyAlaGlyAlaThpAaThp 816
QY 393 CTTGATGCTGCGACCGCTTGCGGCACTGACCTGACGCGCTGCGCGCGCGCGCGCGCG 452
DB 816 IAlaLeuAlaThpThpAlaThpAaThpPro-----AaGlyAaIaThpAaThpProThpG 832
QY 453 CGTCAACGCGCGCTGATGCTGCGGCAATTAACGCTGCGCGCGCGCTCATGCGGACTT 512
DB 832 LysAaCysPro-----AaThpAlaProAlaThpAaThpProThpS 845
QY 513 GAGCGCGCGCGCGACCGACGAGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAT 572
DB 845 eAaSerAlaIaThpProPro-----AlaThpProAlaIaIaSerProThpS 860
QY 573 GCGCGACGCGCGCATTCGCGCATTTGCGCGCGCGCGCTTCTACCGCGCGCGCGCGCGCAC 632
DB 860 LysAlaAlaIaThpProSerAaThpThpAaThpProMetProAla-----ProLysAlaIaIa 876
QY 633 CACCGAAGAGATCATCGAGGTGTGCCGCGCGCTGACGCGCGCGCGCGCGGATCATGACCGAC 692
DB 877 His-----AlaThpAlaGlyThp-----LysSer 884
QY 693 CCACATGCGCGACGACGAGCGGACGACATCGT-----GGCGCGC 728
DB 885 ProLysSerAaThpAaThpAaThpGlnAaThpLeuProGlnLeuLeuAaThpAlaLeuGly 904
QY 729 GCTGAGAGAAACCTTCCGCGCATCGG-----CGCGGACTGGAAGCT 767
DB 905 AlaGlyAlaIaThp-----ProAlaThpProGlnProAaThpAaThpLeuValaThpAaThpAla 923
QY 768 GCGCGGTGATGATCTCGACACCAAGGTCTATGGGCGCGCGCAATTCGCGCGCGCGCGCGCA 827
DB 924 AlaProGlyLeuLeuAaThpThpAlaGlyLeuGlyAaP-----GlnProLeuGlyAla 940
QY 828 GAC-----GCTGCC 836
DB 941 AspAaThpProAlaValaProGlyLeuAaPaaThpAaThpProAaThpAlaAlaIaThpAaThpPro 960

```

Db	1310	ArgAlaAlaValProArgSerGlyValAlaArgThrArgProGlyArgGlnProGlu		1322
Qy	1569	GGC-----CCCTCCCTCCGCTCCCAATAGCGCCCAACCCGATATGCT-----		1611
Db	1330	GlyArgAlaProProGlnValAlaAspHisArgArgAlaGlnArgHisArgGlyGly		1349
Qy	1614	CAAGGAAGTGTATGGGCGCGGCGCTGGCCGCCGCGCAAGCCCGGAA-----		1661
Db	1350	AlaGlyArgProGlyGlyMetAspArgGlnLeuProArgGlnProGluLeuHisArgArg		1369
Qy	1662	AAATACCTGCAGACACTGTGCGAGGCGCAAGCGGCATGCGGTGTCGACCTGTCCAAAGC		1721
Db	1370	GlyArgProAlaArgArgProAlaGlyAlaArgHisArgValAlaGlnPro-----		1386
Qy	1722	CGAGCTGGCGCAGATCGCCCTGAGTACAGAAAGCT		1757
Db	1387	-----GlyProAlaArgArgProGlyArgThrAla		1396

sequence 26129, Application US/10366683  
: GENERAL INFORMATION:

APPLICANT: Rubenfield, Marc J  
APPLICANT: NO114nd York

APPLICANT: Deloughery, Craig  
APPLICANT: Bush, David

INVENTOR: bush, David  
; TITLE OF INVENTION: NUCLEIC A  
; TITLE OF INVENTION: NUCLEIC A

FILE REFERENCE: PATH03-04

CURRENT APPLICATION NUMBER: US/10/366,683  
CURRENT FILING DATE: 2003-02-13

PRIOR APPLICATION NUMBER: 09/252,993  
PRIOR FILING DATE: 1999-03-10

NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 36130

LENGTH: 885

ORGANISM: Pseudomonas aeruginosa

US-10-366-683-26129

Alignment Scores:  
Pred. No.:

Score: \_\_\_\_\_  
Percent Correct: \_\_\_\_\_

### Best Local Similarity

Query Match:  
DB:

JS-10-009-782-1

57 2000-01-01

2	2126ALC
3	11111111
4	

3 / LeuargPI

Y 111 GCGGCGCG

56 ----- b

Y 171 CGCCGCCG

69 ArgArgAR

231 CTCGCACA

87 Thyroglobulin

07 484944

201 60000000

291 GCAGGGCCG;  
||

106 -----Arg

351 CGCCACCC

—

Db 120 ArgArgProLeuProProAlaArgProAlaGlnGluArg-----CysArgAla 135  
 QY 411 CTTGCGCCGACTACTGAGCGGCTGGCGGCGCCGCGCCGCT-----CAAGCG 461  
 Db 136 ValArgArgArgAlaGlyArg-----ArgSerSerAlaGlyGlyArgGlyMetGlyArgGle 154  
 QY 462 CGCCTGTATGTGGGCGCATTCACAGCTGCGCGCGGCTCATGGCGGACTTCAGCGCCG 521  
 Db 155 AlaLeuAlaGlnGlnProArgProGluGlyArgArgGlnHisArgAsp-----Arg 171  
 QY 522 CGGCACGAGAGAGAAATCGCGCCAT-----GCGGACCTGGCGCGAGAGAC 569  
 Db 172 ProGlnAlaAlaAlaHisArgAlaHisProTyrLeuArgAlaGlyPro----- 187  
 QY 570 CATGCCAGCGCGCCATCGCATTCCTGACCGCGCTTACCGCGCGCGCGCGCGCG 629  
 Db 188 -----ArgArgArgLeuValGlnProAspAlaProArg 198  
 QY 630 CACCCAGCAAGAGATCATGAGGTGTGCGCGCGCGCTGAGCGCGCATGGCGCATCTACG 689  
 Db 199 His-----AspGluProGlyArgPro----- 205  
 QY 690 CACCCACATGCGCGAGAGCGAGACATCGTGGCGCGCGCTGAGAGAACTTCGCGCAT 749  
 Db 206 -----ArgCysHisArgGlyGlyProAlaGlnProGlyAla 218  
 QY 750 CGCGCGCGAGCT---GAGCTGCGCGGTGTGATCTCGCAGCAACAGTTCATGGCGCGACG 806  
 Db 219 LeuProArgSerValGlyArgProAlaGly-----SerProGlyLeuSerProAla 236  
 QY 807 ---CAATTTCGCGCGCTCGCGAGACGCTGCGCGCTGATCGAGCGCGCAT----- 854  
 Db 237 GlyGlnGlyGlnProArgArgGlnProAlaProAlaGlyCysProArgArgGlnProPro 256  
 QY 855 -----GCGCGCGCGAGAGCTGTC 872  
 Db 257 GlyGlyArgCysProArgProCysArgAlaLeuProGlnProGlyGlnProGlyArgGly 276  
 QY 873 GCTGACCGCGATCCCTACGTGGC-----GCG 899  
 Db 277 AlaAlaArgArgArgArgGlyProValArgArgProPheArgCysThrAlaLeuArg 296  
 QY 900 CTCACACATGCTCAAGCAGAGACCGGCTGCTGCGCGCGCGACCGACCATCAGCTG--- 956  
 Db 297 ArgGlnArgAlaAlaLeuProArgSerValAlaGlnArgAsnProAlaProLeuArg 316  
 QY 957 -----GTCAAGCCCTTCCCGCACTAGCGCG---GCGCGACCGATGGAAGTGC 1004  
 Db 317 AlaAlaThrValAlaAlaGlyProAlaArgGlnArgHisArgProAlaArgLeuVal 336  
 QY 1005 GCGCGAGCGCGGCAATCCAAATCAAGTACAGCTGTCGCGCGCGCGCGCGCGCGCAT 1064  
 Db 337 GlyThrAlaArgProValProProAlaArgHisAla---AlaAlaAlaGlyArgThrSer 355  
 QY 1065 CTAATTGATGATGAGACAGACCGGCGCGCGCAT---CTGCGCGTTCGCGCGCGCAT 1121  
 Db 356 ArgArgProValProArgLeuArgArgGlyArgHisArgProGlyAlaAla----- 372  
 QY 1122 GATCGGCTCGAGCGCGCTGCGCA----- 1145  
 Db 373 GlyArgLeuArgProProAlaAlaGlySerProGlyProGlyThrAlaValAlaGlyPro 392  
 QY 1146 CGAGGAGCGCGCCATCCGCGCGCTGTGGCGACCTTCGCGCGGTGTGGAGATGACCGCGCT 1205  
 Db 393 AlaGlnArgProAlaAlaGlyAlaAspAlaHisLeuAsp---GlyLeuArgLeuArgPro 411  
 QY 1206 GCGCGACTGTGGCT-----GTTCCGCGTGAAGAGCGCGGTATGGAAGTGAACCGCGCT 1259  
 Db 412 AlaArgProGlyProAsnHisValAlaValGluSerArgProGluAlaAspProArgPro 431  
 QY 1260 GACCGCGCGCGCTTGGCGCGCTGCGCGCGCGCGCGCGCATGCAAGCGCGGTACTCGCGCA 1319  
 Db 432 AlaArgArg-----ArgProGlyArgHisAlaGlyAlaArgArgProProAlaAlaArg 449  
 Db 433 AlaArgArg-----ArgProGlyArgHisAlaGlyAlaArgArgProProAlaAlaArg 449

QY 1320 CTTGTGTGTGTGACCGCGCGCATGCT----- 1346  
 Db 450 ProGluArgProAlaProGlyThrGlyAlaCysGlnProAlaMetAlaAlaGlyThrArg 469  
 QY 1347 -----GCCGATACCGC----- 1358  
 Db 470 ProValValAlaAlaProAlaGlyArgArgGlyThrGluProGlyLeuArgProAspPro 489  
 QY 1359 -----CACCTTGACACACC 1373  
 Db 490 AlaGlyTyrGlyProAspArgProAlaSerArgThrAlaGlyTyrLeuAlaGlyPro 509  
 QY 1374 TACGAGCGCGCGCGCGCATTCATCCGTATGACGCAACGCGCGCGGTGCGCAAGA 1433  
 Db 510 ValAlaAlaGlyArgProGlnProPheArgProGlnHisProAlaAspGlyLysGlyArg 529  
 QY 1434 GCAGCGCTTCACCGCGCGCGCATGCGCGCGCGCTGTCGCAAG-----CACGCGCGC 1484  
 Db 530 LeuArgThrSerArgAlaLysAlaLeuProArgArgArgAlaValProHisGlySer 549  
 QY 1485 CTGAGCCG-----GCCAGCCCTTACATCCGCGGTGACAGG 1523  
 Db 550 ArgAlaProGlySerGlyGlyArgGlnAlaGlyProGlyAlaArgArgArgAlaGln 569  
 QY 1524 GCGCGGTGCGCGCGCT---CCCAACCTGGA-----CGCAA 1559  
 Db 570 ArgAlaCysProAlaAlaArgProArgProAspLeuValArgAlaThrAlaLeuArgGln 589  
 QY 1560 CCGCTACATGCGCGCTCCCTC-----CGCTCGCATAGCGCGCGCGCATATCGT 1610  
 Db 590 ProAlaHisArgArgThrLeuValGluArgAlaAlaAlaGlyGlyProAlaGlyArgVal 609  
 QY 1611 GCGCAAGAGAGATGAGCGCGCGCGCTGCGCGCGCGCGCGCAAGCGCGGAAATGACCT 1670  
 Db 610 GlyGlyArgGlnAlaGlyArgGlnProAlaArg-----ProGlyArgArgPro 625  
 QY 1671 GCA-----AGACGTGTGCAAGCGCGCGCGCATGCGCGCTCGACCGCGCGCAAGG 1721  
 Db 626 AlaHisLeuGlnArgArgArgAlaGluSerArgGlyProGlyProGluProAlaGlnGly 645  
 QY 1722 CGAGCTGG-----CCAGATCGC 1739  
 Db 646 GluValGlyGlnPheProAlaGluLeuLeuHisGlyLeuProAlaArgHisProAspArg 665  
 QY 1740 CCT 1742  
 Db 666 Pro 666

RESULT 8  
 US-10-419-128-26129  
 ; Sequence 26129, Application US/10419128  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/10/419,128  
 ; PRIORITY FILING DATE: 2003-04-21  
 ; PRIOR APPLICATION NUMBER: US/09/252,991  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 26129  
 ; LENGTH: 885  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-10-419-128-26129  
 Alignment Scores:

Pred. No.: 5.37e-08 Length: 885  
 Score: 434.00 Matches: 214  
 Percent Similarity: 34.66% Conservative: 29  
 Best Local Similarity: 30.53% Mismatches: 248  
 Query Match: 13.16% Indels: 210  
 DB: 6 Gaps: 37

US-10-009-782-1 (1-1758) x US-10-419-128-26129 (1-885)

```

QY 57 CTTGACCTGCTGCTGGCGGGCGG-----CACCTCATGACGAGCAACACCGCGG 110
DB 37 LeuArgProAlaAspArgGlyLeuGlnHisArgHisArgLeuPheGlnArgPro--- 55
QY 111 GCGGCGCGCGACCTGGCGCGTGGCGGACCGGACCGGCGGCGGCGGCGGCGGCGG 170
DB 56 -----ProGlnArgHisAlaArgProAlaArgValArgValArgPro 68
QY 171 GCGGCGCGCGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 230
DB 69 ArgArgArg-----ProGlnArgHisAlaArgProAlaArgValArgValArgPro 86
QY 231 CTGCAACACCGACGAGCAACACTACCTGCTCAGGCGGCGGCGGCGGCGGCGGCGG 290
DB 87 ThrArgHisProSerGlyProAlaProAlaArgValArgValArgValArgValArg 105
QY 291 GCAAGGCGTACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 350
DB 106 -----ArgHisArgProAlaArgHisArgHisArgHisArgHisArgHisArg 119
QY 351 CGGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 410
DB 120 ArgArgProLeuProProAlaArgProAlaArgHisArgHisArgHisArgHisArg 135
QY 411 CTTGCGGCGGCTACCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 461
DB 136 ValArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 154
QY 462 GCGCTGTATGTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 521
DB 155 AlaLeuAlaGlyGlnProArgProGlnArgValArgHisArgHisArgHisArgHis 171
QY 522 CGGCGCGGCGGAGAAATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 569
DB 172 ProGlnAlaAlaAlaHisArgHisArgHisArgHisArgHisArgHisArgHisArg 187
QY 570 CATGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 629
DB 188 -----ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 630
QY 630 CACCGCGGAGAGATCATCGAGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 689
DB 199 His-----AspGlnProGlnArgPro----- 205
QY 690 CACCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 749
DB 206 -----ArgCysHisArgGlyGlyProAlaArgHisArgHisArgHisArgHisArg 218
QY 750 CGGCGCGGAGAGT-----GGAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 806
DB 219 LeuProArgSerValGlyArgProAlaGly-----SerProGlnArgHisArgHisArg 236
QY 807 ---CAATTCGCGCGCGCGGCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 854
DB 237 GtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 872
QY 855 -----GGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 872
DB 257 GtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 872
QY 873 GCTGAGCGGCTATCCCTAGTGGC-----CGG 899
DB 277 AlaAlaAlaArgArgArgArgArgGlyProValArgArgProPheArgCysThrAlaLeuArg 296

```

```

QY 900 CTCACCATGCTCATGACGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 956
DB 297 ArgGlnArgValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 956
QY 957 -----GTGCAAGCGCTTCCCGCAACTGACGCG-----CGGCGGCGGCGGCGGCGG 1004
DB 317 AlaAlaThrValAlaAlaAlaGlyProAlaArgGlnArgHisArgArgProAlaArg 1064
QY 1005 GCGCGGAGCGCGGCAATTCAGATGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1064
DB 337 GtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 1121
QY 1065 CTACTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1121
DB 356 ArgArgProValArgProGlnArgGlyArgHisArgProGlyAlaAlaAlaAla 1145
QY 1122 GATCGGCTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1145
DB 373 GtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 1145
QY 1146 GAGAGAGCGCGCGCATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1205
DB 393 AlaGlnArgProAlaAlaGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1259
QY 1206 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1259
DB 412 AlaArgProGlyProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1319
QY 1260 GACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1319
DB 432 AlaArgArg-----ArgProGlnArgHisAlaAlaAlaArgArgArgProAlaAla 1346
QY 1320 CCGTGTGTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1346
DB 450 ProGlnArgProAlaProGlyThrGlyAlaCysGlnProAlaAlaAlaAlaAla 1358
QY 1347 -----GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1358
DB 470 ProValValAlaAlaAlaProAlaAlaGlyArgArgThrGlnProGlyLeuArgPro 1373
QY 1359 ----- 1373
DB 490 AlaGlyTyrGlyProAlaArgProAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1373
QY 1374 TACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1433
DB 510 ValAlaAlaAlaGlyArgProGlnProPheArgProGlnHisArgProAlaAla 1433
QY 1434 GCGAGCGGTTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1484
DB 530 LeuArgThrSerArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1484
QY 1485 CTGAGCGCG-----CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1533
DB 550 ArgAlaProGlySerGlyGlyArgGlnAlaGlyProGlyAlaAlaAlaAlaAla 1533
QY 1524 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1559
DB 570 ArgAlaCysProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1559
QY 1560 CCGGTCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1610
DB 590 ProAlaHisArgArgThrLeuValGlnArgAlaAlaAlaAlaAlaAlaAlaAla 1609
QY 1611 GCGCAAGAGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1670
DB 610 GtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 1670
QY 1671 GCA-----AGACCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1721
DB 626 AlaHisLeuGlnArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 1721
QY 1722 CGAGCTGGG-----CGAATCGC 1739

```

590 AlaSPARGGLyAlaValAlaAla-----GlyLeuProARGProAlaArgIu 606

NUMBER OF SEQ ID NOS: 33142	SEQ ID NO 19446	LENGTH: 937	TYPE: PR	ORGANISM: Pseudomonas aeruginosa	US-10-419-128-19446
Alignment Scores:					
Pred. No.:	5.75e-08	Length:	937		
Score:	433.00	Matches:	234		
Percent Similarity:	43.29%	Conservative:	41		
Best Local Similarity:	29.18%	Mismatches:	257		
Query Match:	13.13%	Indels:	270		
DB:	6	Gaps:	42		
US-10-009-782-1 (1-1758) x US-10-419-128-19446 (1-937)					
QY 6	CCACCTGATCGCGGAGAGAGATTTCATCTCCATCCGATCCAGCCCTTGACCT				
DB 112	ProteinProGlnAArgTThrAspAlaHisArgPro-----GlyThrGlyArgPro				
QY 66	GCTGCTCGCGGCGGCGACCTTCAT-----CGAGCGGCGAGACACCGCGG				
DB 129	---AlaHisProGlnHisProHisLeuPheGlyProGlyArgAlaArgTArgHisProGly				
QY 111	CGCGGCGCGCGACG				
DB 148	Ala-----GlyArgGlyValArgProAlaArgGln-----				
QY 171	CGCGCGCGCGCGACCGCGGTCGA-----CGTGTGCGGCGCGTGTGTGCGCGCGCTT				
DB 158	-----ProGlnHisLeuAlaArgProGlyProGlyGlnGlnArgGlyArg				
QY 225	CATCGACGCGACACCGCGACGACGACAACTACTGCTCAGCGCGTCCGACATGACGCCAA				
DB 173	AspArgProArgHisProHisArgGlnArgValAlaGlnArgGlyAlaSerAspSerArg				
QY 285	GAT-----				
DB 193	GlnArgGlyAlaValProProArgGlyAspArgGlyThrValAspArgLeuProArgPro				
QY 288	-----CTGCGACGCGGTCACACCGGTGCTCAC-----GGCAATTGCGGCAT				
DB 213	GlyProArgGlyGlyGlnGlySerGlyAspHisArgTThrValAlaHisLeuProArg				
QY 330	CACCTGCGCGCGCTGCT				
DB 233	HisProGlyThrGlyAlaThrArgArgProAspArgArgProArgProAlaLeuGly				
QY 381	-----GACGAGAGCGCGCTCTTACCGTTGAGCGCGCTGCGCGCGCTGCGCGCTGGA				
DB 253	GlyThrAlaGlyValArgArgGlyGlyLeuArgAlaThrArgAlaThrGlnGly				
QY 429	CGGCTGCG				
DB 273	ArgLeuProGlnGlnAlaAlaAlaAlaAlaAlaArgGlyArgLeu				
QY 486	GCTGCG				
DB 287	AlaGlnGlnAlaHisAlaAla-----ArgGlnArgArgGlyAspArg				
QY 543	-----GGCCAT-----CGCGGACCTGCGCGGAGAGCGCGCGCGCGCGCGCGCG				
DB 302	GlyProGlyHisLeuProAlaAlaProAspGlnArgAlaGlnArgArgArgLeuGln				
QY 582	-----GGCCATCGGCATTTC-----GACCGCGCGCTTCTACCGCGCGCG				
DB 332	LeuArgHisArgSerLeuArgProAlaLeuGlnGlyGlnArgArgArgLeuGlyArg				
QY 620	-----				
DB 342	LeuGlnValArgGlnArgGlnArgGlnArgGlnArgGlnArgGlnArgGlnArgGlnArg				



```

OY 621 -----CGCCGCGCCACCACCGAAGCATCATCGAGTGTG----- 656
DB 362 Alaaspeerglnvalalaargprogluhsargleuuglylthraarglythargleu 381
OY 657 -----CGGCGCGGTGAGCG----- 671
DB 382 HlsprollaalaasparvglnphleuaspalaProlaaargluasprleuProaarg 401
OY 672 -----GCATGC 677
DB 402 G1yleuvalargleuargleuaglyalaglyalaaspargyleuargvalglnProalaVal 421
OY 678 CGGATCTAGCCACCCACATCGCCGACGAGCGGACACATCGTGGCCGCGGTGAGGA 737
DB 422 GlnhlsleuvalaInProaP---Argargargvalalaglyargargargalaglygly 440
OY 738 AACCTTCGCGTGG-----CGCGAGCTGAGAGTGGCGGTGTGTAT--- 779
DB 441 G1Y-----HlsarglaValaHlsargglyProaargalaglyarglyleuaspala 458
OY 780 -----CTCGACACCAAGGTCA 797
DB 459 GlnlaalaProlaaIleProlaaasphlsargargalaglyleuaserlaaglnlyal 478
OY 798 GGGCGACCCCAATTTGGCGCGCTCGCGGACGCTGCGGTGATCGAGCGCCATGCG 857
DB 479 Aspproaarglaaleuenglnaarglaalaglythralaglnlaasproaargargpro 498
OY 858 GCGCCAGGAGCTGCTGCTGAGCGCGGTACCTACGTGGCGCGGTCTCCACATGCTCAAGA 917
DB 499 CysProaProaargleuProaIleuargserProaIleuarglnlaProaargpro 518
OY 918 GGAACCGCGTCTGCTGCGCGGACGACATCATCTGTCAGACGCTTCGCCGAACT 977
DB 519 G1YargleuvalaIaglyargglyalaaleuargalaProaIleuargleu 538
OY 978 GAGCGGCGGACGTGATGAAGTGGCGGCGGCGGCAATCCAGTA----- 1028
DB 539 ProaargargproaIaglyargleuenglnaargargalaglyleuargproaIleuarg 558
OY 1029 -----CGAGCTGTGTCGCGGACGCTGAGCGCGGCGGCAATCCAGTA----- 1079
DB 559 G1YargleuvalaIaglyargglyalaaleuargalaProaIleuargleu 574
OY 1080 CGAACCCGAGCGGACGACATCTGCGGCGGTGCGCGGCGGCAATGATGCTCGGAGCGCT 1139
DB 575 ArgproaargleuvalaIaglnaHlsIyala-----AlaIeuaIargpro 589
OY 1140 GCGCCGACGAGCGGCGGACATCGCGGCTGCGGCGGCAATGATGCTCGGAGCGCT 1193
DB 590 AlaaspargargglyalaValaIala-----G1yleuProaargProaIleuarg 606
OY 1194 -----GGGCACTAATGCGCGGCA-----CCTGGGCTGTTCGCGTGGAGCGGCGT 1241
DB 607 ArgleuenglnaIalaIeuleuargargValaIaglnlyleuProaIleuargHlsIyHls 626
OY 1242 ATGGAATGATGCGCGGCTGACCGCGC-----GCGCTTCGG 1277
DB 627 ProaIleuargproaIleuargargphlsargalaIargHlsIasphlsasprProaIleuarg 646
OY 1278 CCGTGGCGGCGGCGGAGCTGAGCGGCGGTCTTCGCGCA-----CCT 1322
DB 647 Alaglyargalaglu---MetalaIaglyMetvalaHlsIasphlsargalaIaglyPro 665
OY 1323 GGTGGTGTGACCGCGGCGGAGCTGCGCGATAC-----CGCCACCTTCGACACACCTAC 1376
DB 666 AlaglyvalaIargIleuIyProaIleuargleuvalaIaglnleuargIaglnlygly 685
OY 1377 CGAGCGCGCGCGGATCATCTCGGTACGACGAGCGGCGCGGTCTGAGAGA--- 1433
DB 686 AspaIaIargIyIleuHlsIargleuenglnaIalaIeuleuProaIleuargIyIleuarg 705
OY 1434 -----GCAAGCGTTCACCGCGGACGATGCGCGGCGGCTGCTCGGACGCGGCG 1481

```

```

DB 706 ProaasphlsIeulaIalaIyargargproaIalaIargProaIleuargIaglnPro 725
OY 1482 CGGCTAGCGCGCGGCGGACGCTTACATTCGCGGTGAACGGGCGCGGTGCGCGCT 1541
DB 726 Pro---AspproaargproaIleuargIaglnleuargIyIleuvalaIalaIa 741
OY 1542 CCCAACCTTGAGCGCAACCGGTACAT----- 1568
DB 742 -----AspaargargargproaIleuHlsIeuleuHlsProaIleuargIalaIeuleuagly 759
OY 1569 GCGCGCTTCCTCGCGCA----- 1589
DB 760 G1YAspaasphlsIargIaglnlaIargargProaIalaIalaIasprleuargProaIa 779
OY 1590 TACGCGCCCGGACGATGCTGCGGAGGAGTGTGGCGCGCGGTCTGCGCGGCGG 1649
DB 780 AlaglyProaIleuargleuenglnaIargIleuargProaIleuargProaIleuarg 799
OY 1650 CAAGGCGCGGAA-----AATGACCTGCAAGCGGTGCGCA----- 1685
DB 800 ArgargProaIleuargleuvalleuargargProaIleuargProaIleuargHls 819
OY 1686 -----GCGCAGCGCATCGCGGTCTGAC----- 1709
DB 820 CysglnaglyalaValaIargleuargHlsProaIleuargHlsProaIleuargHls 839
OY 1710 -----CCTTCAAGCGGAGCTGCGGCGGCAATCGCGGTGAGAGA 1751
DB 840 AspaIalaIeuglnProaIleuargIaglyIleuargIyIleuargIyIleuargIy 859
OY 1752 GAGCT 1757
DB 860 AspaIa 861

RESULT 11
US-10-366-683-30843
Sequence 30843, Application US/10366683
GENERAL INFORMATION:
APPLICANT: Rudenfeld, Marc J.
APPLICANT: Nolling, Jork
APPLICANT: Delounghey, Craly
APPLICANT: Bush, David
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: PAT03-04
CURRENT APPLICATION NUMBER: US/10/366,683
CURRENT FILING DATE: 2003-02-13
PRIOR APPLICATION NUMBER: 09/252,991
PRIOR FILING DATE: 1999-02-18
NUMBER OF SEQ. ID NOS: 3142
SEQ. ID NO 30843
LENGTH: 663
TYPE: PRN
ORGANISM: Pseudomonas aeruginosa
US-10-366-683-30843

Alignment Scores:
Pred. No.: 6.37e-08
Score: 432.50
Percent Similarity: 33.51%
Best Local Similarity: 27.52%
Query Match: 13.11%
DB: 6
Gaps: 40

US-10-009-782-1 (1-1758) x US-10-366-683-30843 (1-663)
OY 32 CGATGCGCCCAATCCGATTCGACG-----CCTTGACCTGCTGCGGCGGCGGACCC 85
DB 4 ProaIleuargproaIleuargargphlsargalaIargHlsIasphlsasprProaIleuarg 23
OY 86 TCATGACGCGGACGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 145

```

Db 24 TrpSerThrThrThrPro-----ProAlaAspArgSerCylValAlaThrAla 39  
 QY 146 TCGCGCCATCGCGCATCTGTGAGACGGCGCGCACACCGGGTGCAGCTGTGGGCC 205  
 Db 40 ArgValProAla-----CysArg-----ArgValProIleArgCysAlaAsnAla 54  
 QY 206 TGGTGGTGGCGCGCGCTTCATGACTGCACACCCACAGACACAACTACTGCTCAGGC 265  
 Db 55 TrpProGlyArgProMetSerSerThr----- 63  
 QY 266 GTCGGCACATGACGCCCAAGATCTGCGAGGGCGCTGCACCGGTGTGTCAGCGCAATTGCG 325  
 Db 64 -----SerArgArg-----ProGlyTrpSerProAlaValPro 74  
 QY 326 GCATCAGCC-----TGGCGC-----CGCTGG 346  
 Db 75 AlaGlyAlaCysCysTrpIleThrArgAsnValAlaIleArgSerAlaThrAspAlaTrp 94  
 QY 347 CGCAGCCCAACCGCGCGCGCGCGCTGACCTGTCAGCAAGAGCGGCTTACCGTTGCG 406  
 Db 95 SerSerAlaProAlaProAlaAsnCysCysLeuPheProAlaGlyProSerProAla 114  
 QY 407 AGCGCTCGCGCATCTGACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 463  
 Db 115 SerProAlaProAlaAlaCysArgArg-----TrpProArgAlaAlaCysHisTrpProAla 133  
 QY 464 CCGTATGTCGGCGCATCTCAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523  
 Db 134 SerAlaTrpTrpTrpLeu-----AlaProAlaArgCysCysTrpProAlaPro 150  
 QY 524 CCACGACGACGAGAAATCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571  
 Db 151 AlaProAlaSerAlaGlyArgAlaCysCysAlaSerProAsnArgArgArgAlaGluPro 170  
 QY 572 TGGCGACGCGCGCGCATCGCGCATTCGACCGCGCGCGCTTCTACG-----CGCGCGCGCGCG 625  
 Db 171 TrpProSerProTyrAlaSerArgAlaGlyProAlaSerCysGlyArgTrpProAla 190  
 QY 626 ----- 658  
 Db 191 CysSerProValAlaThrAlaProThrAlaThrCysSerProProSerAlaArgSerAla 210  
 QY 650 -----AGGTGTGCC----- 658  
 Db 211 TrpLysProCysAlaCysAlaLeuAlaValAlaIleYserAlaArgSerProAlaAsnAlaTrp 230  
 QY 658 ----- 658  
 Db 231 ProAlaAlaSerAlaTrpCysProThrCysAsnSerAlaArgProSerAlaIleAlaSer 250  
 QY 659 -----GGCGCGTGAAGCGCGCATGGCGCGCATCTACGCCA 691  
 Db 251 ThrAlaAlaArgSerProSerThrAsnGlyArgProAlaAlaCysProThrThrThrPro 270  
 QY 692 CCCACATAGCGCG-----ACGAAGCGGACGACATCGTGGCGCGCGCGGAGAAACCTTCGCGA 748  
 Db 271 ProAlaAsnAlaProAlaSerAlaAlaAlaAsnTrpTrpTrpLysAlaProSerPro 290  
 QY 749 TCGCGCGCGAGTGGACGTGGCGGTGTGATCTCCGACCAACAGTCAATGGCGCGCGCGCA 808  
 Db 291 AlaMetProProSerThr----- 296  
 QY 809 ATTTGGCGCGCTCGCGCGAGACGCTGC-----CGCTGATCAGAG 847  
 Db 297 -----SerAlaThrArgLeuAlaAlaCysGlyHisAspGlyAspAlaGlyArgAlaSerGln 315  
 QY 848 CCGCCATGGCGC----- 859  
 Db 316 ThrProTrpArgAspThrSerProCysValProAsnCysValAsnTrpArgArgArgGlyLe 335  
 QY 860 -----GCCAGAGCGTCTCGCTGACGCGCTATCCTTACGTGGCGCGCGCTCA 904  
 Db 336 ProTrpSerAlaValAlaAlaLysThrSerProTrp-----ProArgTrpProGlyMet 352

QY 905 CCATGCTCAACAGACCGCGTGC-----TGTGGCGCGGACGACCACTCAATCA 952  
 Db 353 ProAlaGlyProArgProSerCysThrAlaValAlaAlaTrp-----AlaProAlaArg 370  
 QY 953 CCGTGTGCAAGCCCTTCCCGCAACTGACGGCGCGCGACCTGATGATGGCGCGCGAGC 1012  
 Db 371 GlyIleTyrAlaAlaAlaAlaProArg-----AsnSerCys-----SerAlaGlyArgLe 386  
 QY 1013 GCGGCAATTCAGATGACGAGTGTGGCCCACTGCA-----GCCGGCG 1057  
 Db 386 ValArgArg-----ArgValArgArgSerAlaArgAlaTrpArgProTrpProAlaGlyArg 405  
 QY 1058 GCGCCATCTACTCATGATGAGACGACCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1111  
 Db 405 GAlaThrProAlaSerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1111  
 QY 1112 GCCGACCATGATCGGCTCGACGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1171  
 Db 425 AlaAlaAspAlaProAlaAlaAlaTrpValSerAlaAlaAlaArgArgThr-----SerSerAlaProIle 444  
 QY 1172 G-----GGGACCTTCGCGCGCGGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1214  
 Db 444 AlaGlySerAlaProGlySerGlyThrAlaProArgCysHisProValArgLysAspGly 464  
 QY 1215 -----GGGCGCTG-----TCCGCTGACGACGCGCGGTATGAGATGACGCGCGCG 1258  
 Db 464 ValAlaGlyProAlaSerThrGlyArgSerArgArgArgTrpAlaAlaAlaProAlaAlaArg 484  
 QY 1259 ----- 1267  
 Db 484 GAlaArgThrGlyIleArgArgThrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1267  
 QY 1268 -----CGGCGTTCG 1293  
 Db 504 ValArgThrProAlaAlaAlaArgArgArgSerAlaAlaArgThrAlaProArgProAlaIle 524  
 QY 1294 CAGCTGACGCGCGCGGTACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1353  
 Db 524 AlaHisArgProSerAlaSerAlaSerAlaThrGlyTrp-----ProGlyProAlaLeuAlaCys 542  
 QY 1354 ACCGCGACCTTGACACACCTTACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413  
 Db 542 sProAlaAlaAlaGlyArgArgProSerProAlaProAlaProAlaProAlaProAla 562  
 QY 1414 GCGCGCGCGGTGCGCAAGCAGG-----CGTTCACCGCGCGCGCGCGCGCGCGCGCG 1467  
 Db 562 ArgProAlaAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 582  
 QY 1468 CTCGACGACGACGCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1527  
 Db 582 AlaAlaProAlaGlyProThrProAlaArgProPro-----AlaAlaAlaThrGlyArg 600  
 QY 1528 GCGTGGCG 1587  
 Db 600 gProSerAlaProAlaArgProAlaAlaArgArgArgArgArgArgArgArgArgArgArg 620  
 QY 1588 AATA-----CGGCGCGCAACCGCATTCGTGGCGCAAGAGTATGGCGCGCGCGCG 1638  
 Db 620 gCysSerProArgArgProThrProArgSer----- 630  
 QY 1639 GCGCGCGACGACGACG 1698  
 Db 631 AlaProGlyAlaAlaThrGlyAla-----ProThrThrGlyAlaAlaAlaArgPro 646  
 QY 1699 GCGGTCTGACCCCTGTCCAAAGCGGAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1738  
 Db 647 ArgArgArgArgAlaAlaAlaAlaAlaProGlyLysArgArgSer 659  
 RESULT 12  
 US-10-419-128-30843  
 ; Sequence 30843, Application US/10419128  
 ; GENERAL INFORMATION:

US-10-009-182-1 (1-1758) X US-10-419-128-30843 (1-663)

[illegible]

Db	39	glntrgarglnproalargalaglnalaglnlproglalalalaglnasprhshlsarg	58
OY	213	CGCGCGCGGCTTCACTGCATCGCACACCCACGA-----	24
Db	59	glnproarglnglnlnatrgleuatharghshsproalrglnproglalarghshslurproalpro	78
OY	246	---CGACAACTACCTGCTCAG---GGTCCGCAATGACGCCCAAGATCTCGAGGGGT	29
Db	79	glnargargargproalaglnproalaglnargargproaspleuagln-----Progllyarg	96
OY	300	CACCAACGGTGCATCGGCGCAATTGGCGCATACGCCGGC-----GCCGCT	34
Db	97	Argproargglalaglnargarg-----glnproglglinproproglalhlslalapro	11
OY	345	GGC-----GCAGCGCAACCGCGCGCGCGCGCT	37
Db	114	glnproglalarglnproalalalargtrargtrargalaglnlproglalargpropro	133
OY	372	GGACCTGCTGGAGA-----AGCGGCTCTTACCGTTT	40
Db	134	glnlgnproglalarglnproalargproalaglnarglnproalargalargargarg	15
OY	405	CGAGCGCTTCGCGCGCATCTGCATCGAGCGCTTGGCGGCGCAACCGCGCGCGCGCGCGC	46
Db	154	Argalaglnlnatrglnlglleuhalasprhalaglnhsharghalavalarglnlgnl--	177
OY	465	CTGTATGGTGGGCATTCAACGCTGCGCGCGCGGTGCATGCCGCACTTGCAGCGCGCGC	52
Db	173	-----Argargarghshsproglalalavalalaglnaspr	18
OY	525	CACCGAGAGAAATGCGCGCAT---CGGGACCTGGCGCGAGAACCATGCGCACGG	58
Db	184	serargargalargargalaglnproalaglnproserleuglnalaglnargargln	203
OY	582	CGCGATCGGCATTTGCACCGCGCGCTTACCGCGCGCGCGCGCGCGCGCACACCGAAGA	64
Db	204	prohlsarghshs-----Argargleu-----Arglnpro-----	213
OY	642	GATCATCGAGGTGTCCGCGCGCTGAGCGCGCATGGCGGCATTCACGCCACCAATCGCG	70
Db	214	-----glntrparproleuarglnarglnargproal	223
OY	702	CGACGAGGCGA-----GCACATGCTGGCGCGCTGGA	73
Db	224	glnproargargproalalarghproglinproglalarglglarglserglargarglngly	243
OY	735	GGAAACCTTCGCGATCGCGCGCGCATGGA---CGTCCGCTGGTGCATCGCACACAA	79
Db	244	Argargargalarghshsargleuqlnproglalargarghalaglnlnproleuathargproal	263
OY	792	GGTCAT-----GGCGCCAGCCCAATTCGGCGCGCTCGCGCGAGACGCTGGCGCT	83
Db	264	glnylshargylsharglallalalproalalargargarg	275
OY	840	GATCGAGGCGCCCATGGCGCGCGCGAGAGCT---CTCGCTGGA	87
Db	276	Asprarglnglnlnatrgarglnproalalargproargproglinarglnhshslalagly	295
OY	879	CGCGATTCCTCACTGGCGCGCTCCACCATGCT-----CAACGAGACCGGCTGCT	92
Db	296	Argtrparproglinlgnatrgleuathargargalaglylulargglnlproglalargpro	315
OY	930	GCT-----GGCGGAGCGCAACCATCATCATACCTGGTGCACGCCCTT	96
Db	316	Alalarglglalarglglalargalargalargalarghshsproprohshs-----	330
OY	969	CCCCGAACTAGCGGCGCGCATCTGGATGAGTCGGCGG-----	1007
Db	331	---Arglnlntrhshshsargargproglalarghalaglnasprproalargargargthr	349
OY	1008	-----CGAGCGCGCGCAAAATCCAGATGAGATGGTGGCC	1040
Db	350	Asprhshslntrproalaglnhshsproalarglnatrgmetasprlnglnlnatrgalalala	369

```

OY 1041 CGAGGTGCA-----GCCGGCGCGGCATCTACTCATGATGAGCAACCGACGTGCA 1094
DB 370 SerrArgThrProGluProGlyHisArgHisLeuProProAspLysArgLysAlaAla 389
OY 1095 GCGCAATCTGGCGTTCGCGCCGACCATGATCGGCTC-----CGACGGCTCGCGA---- 1145
DB 390 GlyGlyAlaValLeuHisThrProArgArgLysGluGlnArgThrProAlaAlaGln 409
OY 1146 CGACGAGCGCGCGCATCCCGCTGTGGGCGACCTCCCGCGGTGCTGGCGCACTATGC 1205
DB 410 ArgArgGlnLeuAlaProArgProGluArgLysPro-----Trp 423
OY 1206 GCGGACCTGGCGCTGTTCCTCGCTGGAGCGCGGTATGGAAGATGACCGGCTGACGC 1265
DB 424 GlnArgProArgHisGlnPro-----ArgArgLeuArgProGluArg 437
OY 1266 CGCGCGCTTCGCGCT-----GGCGGG 1286
DB 438 ArgGlnProArgProArgGlnArgArgLysGlnProArgArgArgGlnHisAlaArg 457
OY 1287 GCGGCGGAGCTGCAGCGCGCGTA-----CTTCGCGCACTGTGTGTGT 1331
DB 458 ArgArgGlnProAlaGlnArgProThrAlaAspHisArgArgArgProArgLysGln 477
OY 1332 CGACCGCGCGCGCGTGGCGCATACCGCACTCGAACAACCCCTACCGCGCGCGCGCG 1391
DB 478 Arg-----ArgGlnSerGlnLeuArgHisPro-----GlyArgArgArg 491
OY 1392 CATCATCTCCCTGACGTCAACGCGCGCGCGCTGCGACAGCAGCGCT----- 1442
DB 492 GlnProAlaLeuArgLysArgLysGlnArgProAlaGlnArgValAlaValGlnThrArg 511
OY 1443 ---CACCGCGCAGCA-----TGC----- 1457
DB 512 ArgHisArgProAlaArgArgArgHisGlnProLeuArgGlyCysValGlnProArg 531
OY 1458 -----CGCGCGCTGCTGCGCACGCGCGCGCTGACGCGCGCGCGCGCGCTT 1505
DB 532 GlnAlaGlnLysArgProArgArgCysArgGlnProCysArgGlnProAlaGlnProPhe 551
OY 1506 ACAATCGCGGTGAAGCGCGCGCGCTGCCCGCTCCCAACCTGAGCAACCGCTA 1565
DB 552 ArgGlyAspArgGlnArgGlnArgLysGlnProAlaCysGlnProAspArgGlu----- 569
OY 1566 CATGCGCTCTCCCTCGCTG-----CAATAGCGCGCGCGCATCTGTGGCAAGA 1619
DB 570 -----ProserArgArgArgLysGlnValGlyProAla-----HisArgGluAspGly 586
OY 1620 AGTGATGGCGCGCGCTGCGCGCGCGCGCGCGCGCG----- 1658
DB 587 AlaAsp-----ArgLeuHisArgArgArgGlnArgGlnArgLysGlnArgGlnAlaArg 605
OY 1659 -----GAAATGACCTTCAGACCTGTCCGACGCGCGCGCGCGCGCGCTCGACCTT 1712
DB 606 HisLeuAspHisThrAlaArgLysGlnAspArgGlyHisArgGlnHisGlnHisGlyAla 625
OY 1713 GTCGCAAGC-----CGAGCTGGCGCGCATGCGCTT 1742
DB 626 ThrAlaGlyArgArgArgLysArgHisArgArgHisProGluGlnPro 642

```

```

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21920
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-419-128-21920

Alignment Scores:
Pred. No.: 6,64e-08 Length: 822
Score: 431.50 Matches: 223
Percent Similarity: 35.84% Conservative: 34
Best Local Similarity: 31.10% Mismatches: 233
Query Match: 13.08% Indels: 227
DB: 6 Gaps: 42

US-10-009-782-1 (1-1758) x US-10-419-128-21920 (1-822)
OY 45 CGATTCACAGCCCTTCGACCTGCTGCGGCGCGCACCCCTCATGACGCGACAGAC 104
DB 2 ArgArgProValAspGlnProGlyProAlaArgArgHisProArgArg----- 18
OY 105 CCGGCGCGCGCGCGCGCGA-----CCTGGCGCTGCGCGCGCGCGCGCGCGCG 152
DB 19 ProGlyLysArgArgArgArgGlyAlaProGlyGlnArgArgArgProAlaAlaAlaGlnArg 38
OY 153 CATGCGGAGTCTGTGCGACCGCGCGCGCGCACACCGCGGTGCGAGTGTGCGGCTGTGT 212
DB 39 GlnArgArgLysProArgArgArgGlnAlaGlnProGlyAlaAlaAlaLysArgHisAlaArg 58
OY 213 CGGCGCGCGCTTCATGACGCGCACACCGCA----- 245
DB 59 GlnProArgGlnLysArgLysArgHisProArgGlnProGlyArgHisGlnProAlaPro 78
OY 246 ---CGACAACTACTCTGAG---GCGTCGCGCATGACGCGCGCGCGCGCGCGCT 299
DB 79 GlyArgArgArgProAlaGlnProAlaGlnArgArgProAspLeuGln-----ProGlyArg 96
OY 300 CACCGAGGTGTGCGCGCGCGCAATGCGCGCATGACGCTGCG-----GCGCGCT 344
DB 97 ArgProArgGlyAlaAlaLysArg-----GlnProGlyGlnProProGlyHisAlaPro 113
OY 345 GGC-----GACGCGCAACCGCGCGCGCGCTT 371
DB 114 GlyProGlyArgGlnProAlaAlaArgArgTrpArgArgLysGlnProGlyArgProPro 133
OY 372 GGACTGTGAGCA-----AGCGCGCTCTTACCGTTT 404
DB 134 GlnGlnProGlyArgGlnProArgProAlaGlnArgGlnProArgGlnArgArgArg 153
OY 405 CGAGCGCTTCGCGCACTTCGACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 464
DB 154 ArgAlaGlnGlnArgGlnGlyLysAlaAspArgLysHisArgAlaValAlaArgGlnGln 172
OY 465 CTGTATGTGGCGCATTAACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 524
DB 173 -----ArgArgArgHisProGlyAlaValAlaAlaLysArg 183
OY 525 CACGACGAGGAAATCGCGCGCAT---GCGGACCTGCGCGCGCGCGCGCGCGCGCGCG 581
DB 184 SerrArgArgAlaArgArgLysGlnProAlaGlnProLeuArgLysGlnArgArgGln 203
OY 582 CGCATGCGCGCTTCGACGCGCGCGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCG 641
DB 204 ProHisArgHis-----ArgArgLeu----- 213
OY 642 GATCATGAGGTGTGCGCGCGCGTGCAGCGCGCATGCGCGCGCGCGCGCGCGCGCG 701
DB 214 -----GlyTrpArgProLeuArgGlnProAla 223

```

[illegible]



```

Db 97 ArgProGly-----GlyGlyArgArgProAlaAlaGlyArgTrpGlnGlyValCys 114
QY 363 CCGCCCCCTGACCTGCTGGAGCGAAGCGGCTTTACCGTTTCGAGCGCTTCGCGACTA 422
   ||| ||||| ||||| |||||
Db 115 ArgThrAspLeuProGlyGlyArg-----ArgArgGln 125
QY 423 CCTGGAGCGCTTGGC-----GGCCAGCGCGCGCGCTCAAGCGCGCTGTATGTTGG 476
   ||||| ||||| ||||| |||||
Db 126 ProGlyAsnArgAlaGlnGlyGlnHisGlyGlyAspArgValAlaArgProGlyAla 145
QY 477 CCAATTCAAGCTGGCGCGCGCTCATGCGCGACTTCGACGG-----518
   ||||| ||||| ||||| |||||
Db 146 -----GlyAlaArgArgGlyArgGlyProAlaGlnAlaProPheArgProArgGln 162
QY 519 -----CGCGCGCGCGAGAGAAATCGCGCGCGCATCGCGGACTTCG 560
   ||| ||||| ||||| |||||
Db 163 ProProAlaLeuProHisArgArgPheArgArgAlaGlyArgAlaAlaSerProAla 182
QY 561 CGAGGAGC-----569
   ||| ||||| ||||| |||||
Db 183 ArgHisAlaAlaHisArgArgProGlyAlaPheLeuProGlyAspAlaAspPheLeuGly 202
QY 570 CAGGCGCAGCGGCGCGCATGCGCATTTGACCGCGGCTTTCACCGCGCGCGCGCGCG 629
   ||| ||||| ||||| ||||| ||||| |||||
Db 203 AsnAlaGlnArgArgArgArgHis--AspArgLeuLeuAspAlaGluProHisArg 221
QY 630 CACCCAGGAGAGATCATCATGAGTGTGCGCGCGCTGACCGCGCATGCGCGCATGACG 689
   ||||| ||||| ||||| |||||
Db 222 SerAlaGlyAlaGlyArgArgGlyProArgGlyPheLeuGlnProHisArgHis 241
QY 690 CACCCACATCGCGG-----CGAGGCGAGCAGCATCGTGGCGCGCT-----731
   ||||| ||||| ||||| |||||
Db 242 AlaProHisGlyArgArgLeuArgArgGlnGlyAsnProGlyArgGlyAlaGlyLeuPro 261
QY 732 -----GAGAGAAACCTTCGCGCGCGCGCGAGCTGACGCTGCGGCTGATGATCGCA 785
   ||||| ||||| ||||| |||||
Db 262 ValArgGlyGlyArgLeuProHis-----ArgAlaSerGlyGlnAspAla 276
QY 786 CCAAGAGCTATGCG-----CCAGCCCAATTTGCGCGCTCGCGCGAGAGAGCT 833
   ||| ||||| ||||| |||||
Db 277 ProAlaAlaHisGlyGlyHisAlaAspHisArgGlnAlaSerProValLeuArgArgIle 296
QY 834 GCGCGCTGATCGAGCGCGCATGCGCGCGCGAGAGCT-----869
   ||| ||||| ||||| |||||
Db 297 ArgArgArgLeuArgArgArgTrpProAlaArgHisProAspArgProGlyArgGln 316
QY 870 -----CTGCGTGGAGCGCGTATCCCTAAGTGGCGCGCTCCACCATGCTCAAGCAGA 920
   ||||| ||||| ||||| |||||
Db 317 LeuArgLeuPheAlaGlyProLeuArgLeuAspArgArgProArgHisValProLeuGly 336
QY 921 CCGCGTGTGCTGGCGCGAGCGACCAT-----CATGAC 953
   ||| ||||| ||||| |||||
Db 337 GlnArgLeuLeuProArgGlnArgHisGlnArgProProLeuGlnAspGlnHisArg 356
QY 954 CTGGTGCAGAACCTTCCCGGAACCTGAGCGCGCGAGCTGATGAATCGCGCGCGAGCG 1013
   ||| ||||| ||||| |||||
Db 357 LeuGlnHisArgLeuProArgPheArgArgAla--ProGly-----AsnGlyArgHis 373
QY 1014 CGGCAAAATCCAGTACGAGCTGTGCCGA-----GCTGCAGCGCGCGCG 1058
   ||| ||||| ||||| |||||
Db 374 ArgGlyAspHisGlyArgArgGlyProGlnProGlyGlnGlySerAlaGlyAlaGln 393
QY 1059 CCGCATCTTCTCATGATGAGAGAACCGAGCGAGCGAGCATCTGCGCTTCGCGCGAGC 1118
   ||||| ||||| ||||| |||||
Db 394 AlaGlnLeuLeuArgGlnGlnArgAlaGlnArg-----HisPro--LeuProPro-- 409
QY 1119 CATGATCGGCTCCGAGCGGCTTCGCGCA-----CGAGAGCGCGCGCATCGCGGCT 1169
   ||||| ||||| ||||| |||||
Db 410 --AspArgArgAlaGlnProValGlyGlyAspAspArgArgAlaGlySerGlnGlnArg 428
QY 1170 GTGGGAGACCTTCCCGCGGCTGCTGGGAGACTA--TGCAGCGAGACTGCGCTGTTCCT 1226
   ||| ||||| ||||| |||||
Db 429 ValArgProProProArgGlyAsnProCysLeuGlnCysArgGlnProGlyPro----- 446

```

```

QY 1227 GCTGAGACGGCGGTATGAAATGACCGGCTGACCGCGCGCGCTTCGCGCGCG 1286
   ||||| ||||| ||||| |||||
Db 447 -----GlnGlnArgPro-----GlyAlaAspProGlyGln 456
QY 1287 GCGGCGCAGCTGCAGCGCGGCTACTTCGCGAGCTGTGCTGTTGACCGCGCGAGCGCT 1346
   ||| ||||| ||||| |||||
Db 457 ValArgHisPheValHisArgAspLeuProGlnArgGlyAlaAspProHisLeu 476
QY 1347 GCGCGATCCGCGCATCTTCGAACACCTTACCGAGCGCGC-----1385
   ||| ||||| ||||| |||||
Db 477 TyrArgArgGlnHisProProGlnProTrpArgHisArgAspGlyProGlyTrpGln 496
QY 1386 -----CGCGCG-----CATCATTCCT 1403
   ||||| ||||| |||||
Db 497 GlnGlyArgProGlyGlyArgArgGlyLeuProGlyArgArgGlyAlaAlaProAspHis 516
QY 1404 GTAGCTCAAGCGCGCGCGCT-----CTGCGAAGAGCAGCGCTT-----CACCGG 1448
   ||| ||||| ||||| |||||
Db 517 ArgHisGlnTrpArgGlnGlyThrGlnTrpLeuAlaHisArgArgLeuGlyHisArg 536
QY 1449 CCAACATGCGCGCGCGCTGTCGACGACGCGCGCGCTGACCGCGCGCGCGCGCTTACA 1508
   ||| ||||| ||||| |||||
Db 537 ProGlnArgGlnGlyArgAlaGlyCysArgArgAsnHisGlnAlaAlaProGly--Gly 555
QY 1509 ATCGCGGCT-----GAACGCGCGCGCTGCGCGCGCGCTCCACACCTGGA 1553
   ||||| ||||| ||||| |||||
Db 556 ValArgArgAlaAlaLeuGlnGlyGlnArgGlyArgArgValProGlnGlnProGly 575
QY 1554 CGCAAAACCGTACATGCGCGCTCC-----1577
   ||| ||||| ||||| |||||
Db 576 AlaArgProArgAlaAspProAlaValArgGlyThrGlyProAlaGlyLeuLeuArgPro 595
QY 1578 -----CTCGCGTCCCAATAC 1592
   ||||| ||||| |||||
Db 596 GlyPheValAlaGlnHisArgLeuLeuProHisAlaGlnAspLeuArgArgProArgAla 615
QY 1593 GCGGCC-----ACCGATATGCTGGGCGAAGATGATGGG 1628
   ||||| ||||| ||||| |||||
Db 616 GlyProTrpProAlaLeuLeuLeuPheCysLeuArgArgArgLeuGlyGlyAspSer 635
QY 1629 CCGCGCGCTGCGCGCGAGGCGGAGGCGGAAATGACCTGCAACACTGTGCGCAGCG 1688
   ||| ||||| ||||| |||||
Db 636 ArgGlyTrpHisArgArgIle-----ProHisAlaAla 646
QY 1689 CAGCGGCAT 1697
   ||||| ||||| |||||
Db 647 HisArgHis 649

```

Search completed: May 11, 2003, 12:28:05  
 Job time : 108 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: May 11, 2003, 12:11:00 ; Search time 232 Seconds

(Without Alignments)  
9771.041 Million cell updates/sec

Title: US-10-009-782-1  
Perfect score: 3299

Sequence: 1 gaattccactgacgcgcga.....ccctgagctacgagaagctt 1758

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 9138288

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp  
-O=/cgn2.1/USPOO/seqs/07052003-122318.23205/app\_query.fasta.1.1927  
-DB=Pending\_Patents\_AA\_Main -QFMT=fastan -SUFFIX=rapm -MINMATCH=0.1 -LOOPEXT=0  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pcr -NOR-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10009782.ecgn.1.1.401 -tunat.07052003.122518.23205 -NCPU=6 -ICPU=3  
-NO\_XLPY -NO\_MAP -LARGUEURY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=10 -THRADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Pending Patents\_AA\_Main:\*  
2: /cgn2.6/ptodata/1/paa/PCrus\_COMB.pcp:\*  
3: /cgn2.6/ptodata/1/paa/US06\_COMB.pcp:\*  
4: /cgn2.6/ptodata/1/paa/US07\_COMB.pcp:\*  
5: /cgn2.6/ptodata/1/paa/US08\_COMB.pcp:\*  
6: /cgn2.6/ptodata/1/paa/US081\_COMB.pcp:\*  
7: /cgn2.6/ptodata/1/paa/US082\_COMB.pcp:\*  
8: /cgn2.6/ptodata/1/paa/US083\_COMB.pcp:\*  
9: /cgn2.6/ptodata/1/paa/US084\_COMB.pcp:\*  
10: /cgn2.6/ptodata/1/paa/US085\_COMB.pcp:\*  
11: /cgn2.6/ptodata/1/paa/US086\_COMB.pcp:\*  
12: /cgn2.6/ptodata/1/paa/US087\_COMB.pcp:\*  
13: /cgn2.6/ptodata/1/paa/US088\_COMB.pcp:\*  
14: /cgn2.6/ptodata/1/paa/US089\_COMB.pcp:\*  
15: /cgn2.6/ptodata/1/paa/US090\_COMB.pcp:\*  
16: /cgn2.6/ptodata/1/paa/US091\_COMB.pcp:\*  
17: /cgn2.6/ptodata/1/paa/US092\_COMB.pcp:\*  
18: /cgn2.6/ptodata/1/paa/US093\_COMB.pcp:\*  
19: /cgn2.6/ptodata/1/paa/US094\_COMB.pcp:\*  
20: /cgn2.6/ptodata/1/paa/US095\_COMB.pcp:\*  
21: /cgn2.6/ptodata/1/paa/US096\_COMB.pcp:\*  
22: /cgn2.6/ptodata/1/paa/US097\_COMB.pcp:\*  
23: /cgn2.6/ptodata/1/paa/US098\_COMB.pcp:\*  
24: /cgn2.6/ptodata/1/paa/US099\_COMB.pcp:\*  
25: /cgn2.6/ptodata/1/paa/US100\_COMB.pcp:\*  
26: /cgn2.6/ptodata/1/paa/US101\_COMB.pcp:\*  
27: /cgn2.6/ptodata/1/paa/US102\_COMB.pcp:\*  
28: /cgn2.6/ptodata/1/paa/US103\_COMB.pcp:\*  
29: /cgn2.6/ptodata/1/paa/US104\_COMB.pcp:\*  
30: /cgn2.6/ptodata/1/paa/US105\_COMB.pcp:\*  
31: /cgn2.6/ptodata/1/paa/US106\_COMB.pcp:\*  
32: /cgn2.6/ptodata/1/paa/US107\_COMB.pcp:\*  
33: /cgn2.6/ptodata/1/paa/US108\_COMB.pcp:\*  
34: /cgn2.6/ptodata/1/paa/US109\_COMB.pcp:\*  
35: /cgn2.6/ptodata/1/paa/US110\_COMB.pcp:\*  
36: /cgn2.6/ptodata/1/paa/US111\_COMB.pcp:\*  
37: /cgn2.6/ptodata/1/paa/US112\_COMB.pcp:\*  
38: /cgn2.6/ptodata/1/paa/US113\_COMB.pcp:\*  
39: /cgn2.6/ptodata/1/paa/US114\_COMB.pcp:\*  
40: /cgn2.6/ptodata/1/paa/US115\_COMB.pcp:\*  
41: /cgn2.6/ptodata/1/paa/US116\_COMB.pcp:\*  
42: /cgn2.6/ptodata/1/paa/US117\_COMB.pcp:\*  
43: /cgn2.6/ptodata/1/paa/US118\_COMB.pcp:\*  
44: /cgn2.6/ptodata/1/paa/US119\_COMB.pcp:\*  
45: /cgn2.6/ptodata/1/paa/US120\_COMB.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2511	76.1	484	24	US-10-009-782-2
2	2481	75.2	484	22	US-09-807-788-4
3	1115.5	33.8	482	18	US-09-489-039A-9018
4	612	18.6	558	21	US-09-770-517C-2
5	453	13.8	2294	16	US-09-252-991A-17231
6	434	13.2	885	16	US-09-252-991A-17231
7	433	13.1	937	16	US-09-252-991A-17231
8	432.5	13.1	663	16	US-09-252-991A-17231
9	431.5	13.1	822	16	US-09-252-991A-17231
10	430.5	13.0	651	16	US-09-252-991A-17231
11	424.5	12.9	1002	16	US-09-252-991A-17231
12	424.5	12.9	1149	16	US-09-252-991A-17231
13	423	12.8	801	16	US-09-252-991A-17231
14	421.5	12.8	863	16	US-09-252-991A-17231
15	413.5	12.5	774	16	US-09-252-991A-17231
16	410	12.4	1706	16	US-09-252-991A-17231
17	408.5	12.4	638	16	US-09-252-991A-17231
18	407.5	12.4	672	16	US-09-252-991A-17231
19	407	12.3	639	16	US-09-252-991A-17231
20	404	12.2	1053	16	US-09-252-991A-17231
21	402.5	12.2	1418	16	US-09-252-991A-17231
22	401.5	12.2	1639	16	US-09-252-991A-17231
23	399	12.1	724	16	US-09-252-991A-17231
24	398.5	12.1	1213	24	US-10-020-359B-95
25	398.5	12.1	802	16	US-09-252-991A-17231
26	395	12.0	631	16	US-09-252-991A-17231
27	394	11.9	719	16	US-09-252-991A-17231
28	394	11.9	782	16	US-09-252-991A-17231
29	392.5	11.9	1225	16	US-09-252-991A-17231
30	392.5	11.9	782	16	US-09-252-991A-17231
31	392	11.9	720	16	US-09-252-991A-17231
32	392	11.9	1228	16	US-09-252-991A-17231
33	391	11.9	757	16	US-09-252-991A-17231
34	390	11.8	977	16	US-09-252-991A-17231
35	388.5	11.8	720	16	US-09-252-991A-17231
36	388	11.8	681	16	US-09-252-991A-17231
37	386	11.7	728	16	US-09-252-991A-17231
38	386	11.7	1706	16	US-09-252-991A-17231
39	385	11.7	762	16	US-09-252-991A-17231
40	384.5	11.9	720	16	US-09-252-991A-17231
41	384	11.6	478	16	US-09-252-991A-17231
42	383.5	11.6	611	16	US-09-252-991A-17231
43	383	11.6	830	16	US-09-252-991A-17231
44	382.5	11.6	671	16	US-09-252-991A-17231
45	382.5	11.6	701	16	US-09-252-991A-17231

#### ALIGNMENTS

RESULT 1  
US-10-009-782-2  
Sequence 2, Application US/10009782  
GENERAL INFORMATION:  
APPLICANT: TAKEDCHI, Ken-ichi  
APPLICANT: KOIDE, Yoshinao  
APPLICANT: HIROSE, Yoshiniko  
APPLICANT: MORIGUCHI, Mitsunori  
APPLICANT: ISOSE, Kimiyasu  
TITLE OF INVENTION: TRANSFORMED MICROORGANISM AND PROCESS FOR PRODUCING D-AMINOAC.  
FILE REFERENCE: 217301USOCP  
CURRENT APPLICATION NUMBER: US/10/009,782  
PRIOR FILING DATE: 2001-12-17  
PRIOR APPLICATION NUMBER: PCT/JPO0/03932  
PRIOR FILING DATE: 2000-06-15

NUMBER OF SEQ ID NOS: 2  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 2  
 LENGTH: 484  
 TYPE: PRT  
 ORGANISM: Alcalligenes xylosoxydans subsp. xylosoxydans  
 US-10-009-782-2

## Alignment Scores:

Pred. No.:	7.05e-134	Length:	484
Score:	2511.00	Matches:	484
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	76.11%	Indels:	0
DB:	24	Gaps:	0

US-10-009-782-1 (1-1758) x US-10-009-782-2 (1-484)

```

QY 34 ATGCCCATCCGATCCGACCCCTGCTGCTGCGGGGCGGACCCCTCATCGAC
DB 1 MetSerGlnSerAspSerGlnProPheAspLeuLeuAlaGlyThrLeuIleAsp 20
QY 94 GGCAGCAACACCCGGGGGGGGCGGCGGACCTGGGCGGCGGCGGACCGCATCGCGCC
DB 21 GlySerAsnThrProGlyArgAlaArgAlaAspLeuGlyValArgGlyAspArgIleAlaAla 40
QY 154 ATCGCGCATCTGTGCGAGCGCGCGCGCGACACCGGGTGCAGCTGTGCGGGCTGGTGC
DB 41 IleGlyAspLeuSerAspAlaAlaAlaIleThrArgValAspValSerGlyLeuValVal 60
QY 214 GCGCGCGGCTTCATCGACTCGCGACACCCGACGACGACGACGACGACGACGACGAC
DB 61 AlaProGlyPheIleAspSerHisThrHisAspAsnThrLeuAlaArgAlaAsp 80
QY 274 ATGACGCCCAAGATCTGCGAGGGGCGTCAACGCGTGCACGGGCGCATTTGGGCGATCAGC
DB 81 MetThrProIleSerGlnGlyValThrValAlaThrGlyAsnGlyValSer 100
QY 334 CTGGCGCGGCTGGCGGCGGCGGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC
DB 101 LeuAlaProLeuAlaHisAlaAsnProProAlaProLeuAspLeuAspGlnGly 120
QY 394 TCTTACCGTTTCGAGCGCTTCGCGGACGACCTGACGCGCGCTTGGCGGCGCGCGCGCGC
DB 121 SerTyrArgPheGlnArgPheAlaAspTyrLeuAspAlaLeuAlaThrProAlaAla 140
QY 454 GTCACGCGCGCTGTATGGTGGGCGATCAACGCTGCGCGCGCGCGCGCGCGCGCGCGCGC
DB 141 ValAsnAlaAlaCysMetValGlyHisSerThrLeuArgAlaAlaValMetProAspLeu 160
QY 514 CAGCGCGCGCGCGCGCGGAGGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC
DB 161 GlnArgAlaAlaThrAspGlnGlyIleAlaAlaMetArgAspLeuAlaGlnGlnAlaMet 180
QY 574 GCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC
DB 181 AlaSerGlyAlaIleGlyIleSerThrGlyAlaPheTyrProProAlaAlaArgAlaThr 200
QY 634 ACCGAGAGATCATGAGGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC
DB 201 ThrGlnGlnIleIleGlnValCysArgProLeuSerAlaIleGlyIleTyrAlaThr 220
QY 694 CACATCGCGGAGAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC
DB 221 HisMetArgAspGlnGlyGlnHisIleValAlaAlaLeuGlnIleThrPheArgIleGly 240
QY 754 CCGGACGCTGAGCTGCGGGGCGGATCTGCGACGACGACGACGACGACGACGACGACGAC
DB 241 ArgGlnLeuAspAlaProValAlaIleSerHisHisIleValAlaMetGlyGlnProAsp 260
QY 814 GCGCGCGTGGCGGAGAGCGTGCAGTGCAGTGCAGGCGCGCGCGCGCGCGCGCGCGCGC
DB 261 GlyArgSerArgIleThrLeuProLeuIleGlnAlaAlaMetAlaThrGlnAspValSer 280
  
```

```

QY 874 CTGACGCGTATCCCTACGTGGCGGCTCCACCATGCTCAACAGGACCGCGTGTGCTG
DB 281 LeuAspAlaTyrProTyrValAlaGlySerThrMetLeuGlnAspArgValLeuLeu 300
QY 934 GCGGAGCCGACCATATATACCGTGTGCAAGCCCTTCCCGAATGAGCGGCGGCGCGCTG
DB 301 AlaGlyArgThrIleIleThrIleThrPheCysArgProPheProGlnLeuSerGlyArgAspLeu 320
QY 994 GATGAGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB 321 AspGlnValAlaAlaGlnArgGlyGlySerIleTyrAspValAlaProGlnLeuGlnPro 340
QY 1054 GCGGCGCGCATCTTACTTATGATGAGCAACCGGACGAGCGGCGGCGGCGGCGGCGG
DB 341 AlaGlyAlaIleTyrPheMetMetAspIleProAspValGlnArgIleLeuAlaPheGly 360
QY 1114 CCGACCATGATCGGCTCGGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB 361 ProThrMetIleGlySerAspGlyLeuProHisAspGlnArgProHisProArgLeuThr 380
QY 1174 GGCACCTTCGCGGCGGCTGCTGGGCGACTATGCGGCGGCGGCGGCGGCGGCGGCGG
DB 381 GlyThrPheProArgValLeuGlnHisTyrAlaArgAspLeuGlyLeuPheProLeuGln 400
QY 1234 ACGGCGGTATGAGATGATACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB 401 ThrAlaValThrPheMetThrGlyLeuThrAlaAlaArgPheGlyLeuAlaGlyArgGly 420
QY 1294 CAGCTGAGGCGGCGGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB 421 GlnLeuGlnAlaGlyTyrPheAlaAspLeuValAlaPheAspProAlaThrValAlaAsp 440
QY 1354 ACCGCGCATCTTGAACACCTTACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB 441 ThrAlaThrPheGlnHisProThrGlnArgAlaAlaGlyIleHisSerValTyrValAsn 460
QY 1414 GCGCGCGGCGGCTGGGCGGAGAGAGCGGCTTACCGGCGGCGGCGGCGGCGGCGGCGG
DB 461 GlyAlaProValThrGlnGlnGlnAlaPheThrGlyGlnHisAlaGlyArgValLeuAla 480
QY 1474 CCGACGCGCGCGC 1485
DB 481 ArgThrAlaAla 484
  
```

## RESULT 2

US-09-807-788-4

Sequence 4, Application US/09807788

GENERAL INFORMATION:

APPLICANT: Stephen John Clifford Taylor

APPLICANT: Robert Christopher Brown

TITLE OF INVENTION: AMINOACYLASE AND ITS USE IN THE PRODUCTION OF D-AMINOACIDS

FILE REFERENCE: GJE-239

CURRENT APPLICATION NUMBER: US/09/807,788

CURRENT FILING DATE: 2001-04-18

PRIOR APPLICATION NUMBER: PC7/GB99/03458

PRIOR FILING DATE: 1999-10-20

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patent Ver. 2.1

SEQ ID NO 4

LENGTH: 484

TYPE: PRT

ORGANISM: Alcalligenes

US-09-807-788-4

## Alignment Scores:

Pred. No.:	3.53e-132	Length:	484
Score:	2481.00	Matches:	478
Percent Similarity:	99.38%	Conservative:	3
Best Local Similarity:	99.76%	Mismatches:	3
Query Match:	75.20%	Indels:	0
DB:	22	Gaps:	0

US-10-009-782-1 (1-1758) x US-09-807-788-4 (1-484)

QY 34 ATGTCCCAATCCGATTCAGCCCTGACGCTGCTGCGGGGCGGACCTCATACAC 93  
 Db 1 MetAlaGluSerAspSerGlnProPheAspLeuLeuValGlyGlyThrLeuLeuAsp 20  
 QY 94 GGCAGCAACACCCCGGGGCGCGCCGACCTGGCGGTGGCGGGGACCGCATCGCGCG 153  
 Db 21 GlySerAsnThrProGlyAlaGlyAlaAspLeuGlyValAlaGlyAspArgIleAlaAla 40  
 QY 154 ATCGGCATCTGTGGAGCGCGCGCGCACACCGGCTGACGCTGCGCGCTGGTGTG 213  
 Db 41 IleGlyAspLeuSerAspAlaAlaAlaHisThrArgValAspValSerGlyLeuValVal 60  
 QY 214 GCGCGCGCTCATGACGTGACACCCAGACGACGACACTACTCTGACGCTGCGAC 273  
 Db 61 AlaProGlyPheIleAspSerHisThrHisAspAspAsnTyrLeuLeuArgArgAsp 80  
 QY 274 ATGACGCCCAAGATCTGCGACGGGCTCACACGCTGTACAGGGCAATTGGGGCATCAGC 333  
 Db 81 MetThrProGlyIleSerGlnGlyValThrValValThrGlyAspGlyIleSer 100  
 QY 334 CTGGCGCGCTGGCGGCGACGCGACCCCGCGCGCGCGCGCGCTGACGCGGACGCGCG 393  
 Db 101 LeuAlaProLeuAlaHisAlaAspProProAlaProLeuAspLeuAspGlyGly 120  
 QY 394 TCTTACCGTTTCAGAGCGCTTCCGCGGCTGACCTGACGCGGCGCGCGCGCGCGCG 453  
 Db 121 SerTyrArgPheGlnAlaGlyPheAlaAspTyrLeuAspAlaLeuArgAlaThrProAlaAla 140  
 QY 454 GTCAACGCCGCTGTANTAGTGGGCTTCAACGCTGCGCGCGCGCGCGCGCATGCGAGCTG 513  
 Db 141 ValAsnAlaAlaCysMetValGlyHisSerThrLeuArgAlaAlaValMetProAspLeu 160  
 QY 514 CAGCG 573  
 Db 161 GlnArgAlaAlaThrAspGlnGlnIleAlaAlaMetArgAspLeuAlaGlnAlaMet 180  
 QY 574 GCGAGCG 633  
 Db 181 AlaSerGlyAlaIleGlyIleSerThrGlyAlaPheTyrProProAlaAlaGlyAlaThr 200  
 QY 634 ACCGAAGAGATCATGAGGTGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCATCTACGCGCAC 693  
 Db 201 ThrGlnGlnIleIleGlnValCysArgProLeuSerAlaHisGlyIleTyrAlaThr 220  
 QY 694 CACATCG 753  
 Db 221 HisMetArgAspGlnArgGlnHisIleValAlaAlaLeuGlnIleThrPheArgIleArg 240  
 QY 754 CCGGAGCTGAGAGCTGCGGCGGTGATCTGCGACACAGGTCTATGGCGCGCGCGCGCGCGCG 813  
 Db 241 ArgGlyLeuAspValProValIleSerHisHisIleValMetGlyGlnProAspPhe 260  
 QY 814 GCGCGCTCGCGGAGAGCGGTGCGGTGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873  
 Db 261 GlyArgSerArgGlnThrLeuProLeuIleGlnAlaAlaMetAlaArgGlnAspValSer 280  
 QY 874 CTGGAGCGCTATCCCTACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933  
 Db 281 LeuAspAlaTyrProTyrValAlaGlySerThrMetLeuGlnAspArgValIleLeuLeu 300  
 QY 934 GCGGAGCG 993  
 Db 301 AlaGlyArgThrIleIleThrTyrPcysLysProPheProGlnLeuSerIleArgAspLeu 320  
 QY 994 GATGAAGTCG 1053  
 Db 321 AspGlnValAlaAlaGlnArgGlyLysSerLysTyrAspValValProGlnLeuGlnPro 340  
 QY 1054 GCG 1113  
 Db 341 AlaGlyAlaIleTyrPheMetMetAspGlnProAspValGlnArgIleLeuAlaPheGly 360

QY 1114 CGGACCATGATCGGCTCCGACGGCTCGCGGCGACGAGCGCGCGCGCATCGCGCGCTGG 1173  
 Db 361 ProThrMetIleGlySerAspGlyLeuProHisAspGlnAlaGlnProHisPheArgLeuThr 380  
 QY 1174 GGCACCTTCCCGCGGCTCGGGGCGACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1233  
 Db 381 GlyThrPheProArgValLeuGlyHisTyrAlaArgAspLeuGlyLeuPheProLeuGln 400  
 QY 1234 ACGGCGTATGAGATGACGCGGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1293  
 Db 401 ThrAlaValThrPheGlnHisProThrGlnArgAlaAlaGlyIleHisSerValTyrValAsn 420  
 QY 1294 CAGCTGACGGCGGCTGCTTTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 1353  
 Db 421 GlnLeuGlnAlaGlyTyrPheAlaAspLeuValAlaPheAspProAlaThrValAlaAsp 440  
 QY 1354 ACCGCGACCTTCCGAAACCCCTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413  
 Db 441 ThrAlaThrPheGlnHisProThrGlnArgAlaAlaGlyIleHisSerValTyrValAsn 460  
 QY 1414 GCGCGCGCGCTGTGGAAGAGCAGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1473  
 Db 461 GlyAlaProValThrPcysGlnGlnAlaPheThrGlyGlnHisAlaGlyArgValLeuAla 480  
 QY 1474 CGCAGCGCGCGCG 1485  
 Db 481 ArgThrAlaAla 484

RESULT 3  
 US-09-489-039A-9018  
 ; Sequence 9018, Application US/09489039A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709.2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 9018  
 ; LENGTH: 492  
 ; TYPE: PRN  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-9018

Alignment Scores:  
 Pred. No.: 7.8e-55  
 Score: 115.50  
 Percent Similarity: 62.708  
 Best Local Similarity: 48.168  
 Query Match: 33.818  
 DB: 18 Gaps: 4

US-10-009-782-1 (1-1758) x US-09-489-039A-9018 (1-492).  
 QY 19 GAAGGAGAGATTCATGTCATCCCAATCCGATCCCAAGCCTTGACCTGCTGCTCGCGGCG 78  
 Db 9 LysGlyValIleIleAlaMetLys-----ValAspTrpLeuPheLysAsn 22  
 QY 79 GGCACCTCATGACG 138  
 Db 23 ValThrValIleAspGlySerGlyGlyProGlnTyrArgAlaAspValAlaValIleGly 42  
 QY 139 GACCGCATGCG 198  
 Db 43 AspArgIleMetAlaIleAlaProAlaLeuAspValAlaAlaGlnGlnValIleAspGly 62  
 QY 199 TCGGCGCTGCTGCTGCG 258  
 Db 63 GlnGlyArgValLeuAlaProGlyPheIleAspValHisThrHisAspAspIleAsnVal 82

QY 259 CTCAGCGCTGCGACATGACGCCCAAGATCTCGAGGGGCTGACACCGGTGCTACAGGCG 318  
 Db 83 IIEATGMEPRGILUTYRLEUPROLYSLEUSERGLNGLYALVTHRTYALILEVALGILY 102  
 QY 319 AATTGGCGCATAGCCCTGGCGCGCTGGCG---CACGCCAACCCGCCGCCCTGGAGC 375  
 Db 103 ASNCYSGLYILESETRALALATHRALATHMETARGLYGLVALPRASPRROMETHSN 122  
 QY 376 CTGGCTGAGAGGGGGCTTACCGCTTCAGACCGCTTCGCCGACTACCTGAGCCGTTG 435  
 Db 123 LEULENGLYGLNGLGNLHNSPHELETRPRORHVALGLUALATYALAHNSALVAL 142  
 QY 436 CGGGCCACGCGCGCGCGCTCAACGCCCTGTATGTGGCGCATTCAGCGCTGGCGCGC 495  
 Db 143 GLUALALARGPROSETRLEUANSVALGLYTHRTLEULIEGLYHNSITHALAEUARGASN 162  
 QY 486 GCGGTCATCGCGCATTCGACAGCGCGCGCCACCGACAGAGAAATCGCGCGCGCGAGC 555  
 Db 163 ASNHSMETASPRLEUPHETPRORALASNGILUTHTGILILEALIGLYMETARGVAL 182  
 QY 556 CTGGCGGAGAGAGCGCGCGCATCGCGCATTCGACCGCGCGCGCTTCACCGC 615  
 Db 183 GLNLEUARGSPRALAEUARGGLNGLYALAEUGLYLEUSERTHTGILYEUALATYALA 202  
 QY 616 CCGCGCGCGCGCGCGCACCGACAGATCATCGAGTGTGCGCGCGCGCTGAGCGCGCAT 675  
 Db 203 SERALPHENGLINERTHTHTGILUGLVALMETALAEUAGLUGLLEUVALAEGLY 222  
 QY 676 GCGGCGCATTCAGCGCACCCACATCGCGCGACAGAGCGAGACATCGTGGCGCGCTGAG 735  
 Db 223 LYSGLYVALTYRTHTHNSLEUARGSERGLUPHEGLUPROILEUGLVALAEUASPR 242  
 QY 736 GAAACCTTCGCGATCGCGCGAGAGTGTGCGCGGTGATGTCCACCAAGATC 795  
 Db 243 GLUALAPHEARGILEGLYARGHNSGLYASNVALPROVALVALSERHNSHNSLYSCYS 262  
 QY 796 ATGGCGCCAGCCCAATTCGCGCGCTGCGCGCGAGACGCTCGCGTATGAGCGCGCATG 855  
 Db 263 ALAELALATYASANTPRGLYARGTHRTYSLUTHTLEUALAPHEPHASPRGLUMETARG 282  
 QY 856 GCGGCGCGAGAGTGTGCGTGGAGCGCGATACCGCTACGCGCGCGCTGACAGTGTCAAG 915  
 Db 283 GLNGLNGLASPRILEALACYSASPCYSYRPROTYRSEALASERSETHLEUASPR 302  
 QY 916 CAGGACGCGCTGTGCTGGCGCGAGCACATCATCATCTGATCGAAGCCTTCCCGGAA 975  
 Db 303 METLYSGINVALTHNARGLUPHEASPRILEVALILETHRTIPRSEGLUALAGLNPBGILN 322  
 QY 976 CTGAGCGCGCGCGCATGATGAAGTGGCGCGCGCGCGCAATCCAACTAGTACAGCTG 1035  
 Db 323 GLNLAELLYSTHTLEUGLNGLNLLEALASPRGLUTPRGLNVALSERLEUHSASPRALA 342  
 QY 1036 GTGGCGCGAGCGTGGCGCGCGCGCGCATCTACTTCATGATGACAGAAACCGAGTGCAG 1095  
 Db 343 ALAALARGLEUWESTPRORALAGLYALALEYTHNSASMETASPRGLUGLNASPRVALARG 362  
 QY 1096 CGCATCTGCGGTGCGCGCGCGCATGATCGCGTCCGACGCGCGTGGCGCGAGCAGCGC 1155  
 Db 363 ARGVALMETARGYRPRORVALTHNMETILEGLYSERASPRGLYEUAPROBANSAPROMET 382  
 QY 1156 CCGGATCGCGCGCTGTGGGCGACCTTCCCGCGGTGCTGGGCGCATTCGCGCGAGCTG 1215  
 Db 383 PROHISPRARGLEUTPRGLYALAPHEPRORARGVALLEUULYHNSYRTERALYDARGGLN 402  
 QY 1216 GCGCTGTTCGCGTGGAGAGCGCGGTGTGAAGATGACGCGCGCTGACGCGCGCGCTTC 1275  
 Db 403 GLNLEUPHETPRORLEUTHTHTALAVAHNSLYMETHTHGLYLEUSERTALALARGPHE 422  
 QY 1276 GCGCTGCGCGCGCGCGAGTGGAGCGCGGACTGTGCGACCGCGGTGGTGTGAC 1335  
 Db 423 GLNLEUALASPRARGGLYEUVALYSLIEGLYTRPHEALASPRLEUVALLEUPHEASPR 442

QY 1336 CCGGCGAGGCTGCGCGCATACCCCACTTCGAAACCTTACGAGCGCGCGCGCATC 1395  
 Db 443 PROGLNTHVALARGASPRVALASERPHESERASPRORALASPRILE 462  
 QY 1396 CATTCGCTGATGACGAGCGCGCGCGCTGG---CAAGACAGCGCTTCACCGCGCG 1452  
 Db 463 GLUALAVALEMETVALASNGLYALMETSERTYRGLYSERASPRLYSLIETHRTGLARG 482  
 QY 1453 CATGCGCGCGCGCGTGTGCGCACGC 1476  
 Db 483 ---ALAGLYARGPHELEUARGARG 489  
 RESULT 4  
 US-09-770-517C-2  
 : Sequence 2, Application US/09770517C  
 : GENERAL INFORMATION:  
 : APPLICANT: Mitsuhashi, Kazuya  
 : APPLICANT: Yamamoto, Hiroaki  
 : APPLICANT: Matsuyama, Akimobu  
 : APPLICANT: Tokuyama, Shinji  
 : TITLE OF INVENTION: D-AMINOACYLASE AND GENE ENCODING THE SAME  
 : FILE REFERENCE: 06501-072001  
 : CURRENT APPLICATION NUMBER: US/09/770, 517C  
 : PRIOR FILING DATE: 2001-01-26  
 : PRIOR APPLICATION NUMBER: JP 2000-019080  
 : PRIOR FILING DATE: 2000-01-27  
 : PRIOR APPLICATION NUMBER: JP 2000-150578  
 : NUMBER OF SEQ ID NOS: 27  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 2  
 : LENGTH: 558  
 : TYPE: PRT  
 : ORGANISM: Hypomyces mycophilus  
 US-09-770-517C-2  
 Alignment Scores:  
 Pred. No.: 2,62e-26  
 Score: 612.00  
 Percent Similarity: 45.32%  
 Best Local Similarity: 31.38%  
 Query Match: 18,55%  
 DB: 21  
 Gaps: 13  
 US-10-009-782-1 (1-1758) x US-09-770-517C-2 (1-558)  
 QY 61 GACCTGCTGCTGCGCGCGCGCGCATTCGACGCGCACACACCCGCGGCGG---CGC 117  
 Db 4 GLUTLEUPHENHNSERATLATHRTVALILETHRTGLYSPRGLUALAGLNPROMETHVAL 23  
 QY 118 GCGGACCTGGGCGTGGCGCGCGCGCATCGCGCATTCGCGCATCTGTG---GAC 171  
 Db 24 ALASPRVALLEUVALSERLYSGLYLEULILEALYSLIEGLYASNPROGLYSERLEASN 43  
 QY 172 GCGCGCGCGACACCGCG---GTGAGCTGTGCGCGCGCTGTGCTGCGCGCGCGCTTATC 228  
 Db 44 ALATHPRASPRHTARGHNSLEUASPRVALTHRTYRTHRTLEUSERTPROGLYPHEILE 63  
 QY 229 GACTCGACACCCAGACGACAACTACTCTGCTGAGCGCGTGGCGCATGAGCCCAAGATC 288  
 Db 64 ASPMETHSLAHNSERASPRLEUTYRLEUENSERHNSPROASPHNSGLUALALYALE 83  
 QY 289 TCGGAGGCGTCAACACGAGTGTGACGCGCAATTCGCGCATGACCGCTGGCGCGCGCG 348  
 Db 84 THRTGLNGLYCYSTHTHTGILVALIGLYGLNASPRLYLSEERTYRTHALAPROILEARG 403  
 QY 349 CAGCGCAACCG---GTGAGCTGTGCGCGCGCGCTGCGCGCGCGCGCTG 372  
 Db 104 ASNVALASPRINLEUARGALALEARGGLNGLNLEALAGLYTRPASNGLYASNPROTHR 123  
 QY 373 GAC-----CTGCTGAGACGAGCGCGCTTCACCGCTTCGAGCGCTTCGCC 417  
 Db 124 ASPGLUGLUCYSARGHTHTHTLEUARGLYVALIGLYMECPHEGLUTPRINTHTHTGLY 143

```

OY 1345 GGGGCGAATACGGCCAGCTTCGTCAACAACCCATACGAGCGGCCGCCCGGCATTCCTGTT 1404
    ||| ||| ::|||::||| ||| ||| ::||| |||
DB 504 VALLVAsMsEtSerThyTrgLUglUProLYsValProSeerArgGLYLearPhenVal 523
OY 1405 TACGTCAACGGCGCGCGGTCTGGCAAGACGAGCGCTTCAACGGCCAGCATGCGCGCCG 1464
    ||||||| ||| ||| ||| ||| |||
DB 524 LeuValAlaSerCylGlnIleAlaValaLarGLUglULysMethrGLYThrArgGLylGLyS 543
OY 1465 GTGTCTGCACGACGACG 1479
    ||| |||::
DB 544 ThrLeuArgArgSer 548

RESULT 5
US-09-252-991A-17231
; Sequence 17231, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17231
; LENGTH: 2294
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17231

Alignment Scores:
Pred. No.: 2,5e-17 Length: 2294
Score: 455.00 Matches: 238
Percent Similarity: 34.51% Conservative: 36
Best Local Similarity: 29.97% Mismatches: 239
Query Match: 13.79% Indels: 281
DB: 16 Gaps: 46

US-10-009-782-1 (1-1758) x US-09-252-991A-17231 (1-2294)
OY 9 CTTGATCGGGAAGGATTTCCATGTCSCAAATCCGATTCCACCACCCCTTGACACTGCT 68
    ||| ||| ::||| ||| ||| |||
DB 675 LeuAlaIarProGlnArgrspAlaArgrProPrheAlaIle-----ProAla 690
OY 69 GCTCGCGGCGGCGCACCTTCATCGACGCGAGCACACCCCGCGGCGCGCGCGACT--- 125
    ||| ||| |||::||| ||| ||| ||| ||| |||
DB 691 AlaValGlnArgTLyArgHisArgProGlyLeuAlaIalGlyLeuAlaValaLarProAla 710
OY 126 -----GGGCGTGGCGGCGGCA-----CCGCATCGCGCGCATCGGCGCATCT 164
    ||| ||| ||| ||| ||| ||| ||| |||
DB 711 AspGluAlaGlnArgTLyArgTLyArgTLyAlaGlnAlaHisProArgHisArgTLyAsp 720
OY 165 GTCGAGGCC----- 173
DB 731 ProGluAlaGlnHisArgTLyArgTLyArgTLyAlaIalAlaGlyAspArgTLyAlaIalThrPro 750
OY 174 -----CGCGCGGACACACCCGGGTGCGAGCGTTCGCGCTGCGTGGTGGCGCGCGG 221
    ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 751 GlyAlaLeuProArgTLyArgTLyArgHisGlnGlnGlyProAlaGlnGlyGlyArgTLyArgTLyArg 770
OY 222 CTTGATCGACTCGGACACCCACGACGACACATACTGCTCAGCGCGCTGCGGACATGACGCG 281
    ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 771 ProLeuGlnGlyAspHisArgTLyArgTLyArgTLyAlaValaLarGlnArgTLyGlnArgTLyAla 790
OY 282 CAAGATCTCGAGGCGCGCACACAGGTGGTCAAGGGGCAATTGCGCATGACGCTGGCGCGC 341
DB 791 -----ArgArgValProAlaIala 796
342 GCTGGCGCA-----CGCAAACCCGCGCGCCCCCTTGACGT---GCTGAGAGAAGGCGG 392

```











```

Db 231 ProAlaAlaSerAlaTrpCysProThrCysAsnSerAlaArgProSerAlaIleAlaSer 250
QY 659 -----GGCGCGTGAAGCGCGATGCGGCGATCTACGCCA 691
Db 251 ThrAlaArgArgSerProSerThrAsnGlyArgProAlaAlaCysProThrThrPro 270
QY 692 CCCACATGCGCG---ACGAAGCGGAGACATCTGGCGCGCGCTGAGAAACCTTCGCA 748
Db 271 ProAlaAsnAlaProAlaSerAlaAlaAlaAsnTrpProThrPylAsnProSerPro 290
QY 749 TCGCGCGGAGCTGAGAGTCCCGGTGATCTCCGACCAAGGTCAAGCGCCACCCCA 808
Db 291 AlaMetProProSerThr-----296
QY 809 ATTTCGCGCGCTGCGGAGACCGCTGC-----CGCTGATCGAG 847
Db 297 ---SerAlaThrArgLeuAlaAlaCysGlyHisAspGlyAspAlaGlyArgAlaSerGln 315
QY 848 CCGGCATGCGCG-----859
Db 316 ThrProThrArgAspThrSerProCysValProAsnGlyValAsnTrpArgArgGly 335
QY 860 -----GCCAGAGCTCCGCTGGAGCGGTATCCCTACGTGCGCGCGCA 904
Db 336 ProThrSerAlaValAlaAlaCysThrSerProThr-----ProArgTrpProGlyMet 352
QY 905 CCATGCTCAAGCAGAGCGCGCTGC-----TGCTGCGCGGAGCGACCATCATCA 952
Db 353 ProAlaGlyProAlaArgProSerCysThrAlaValAlaAlaTrp-----AlaProAlaArg 370
QY 953 CTTGTCGAGAGCGCTCCCGGAACTGAGCGGCGCGGACGATGAGTGGCGCGCGAGC 1012
Db 371 GlyAlaGlyArgAlaAlaAlaProArg-AsnSerCys-----SerAlaGlyArgLe 386
QY 1013 GCGGCAATCCAAATACAGAGTGGTGGCGCGAGCTGCA-----GCCGCGCG 1057
Db 386 ArgArgArg---ArgValaArgArgSerAlaArgAlaTrpArgProThrProAlaGlyArg 405
QY 1058 GCGGCATCTACTCATATGAGAGACACCCGATGCGAGCG-----CATCTGCGCGTTCG 1111
Db 405 GAlaThrProAlaSerAspAlaArgArgAspSerAlaSerArgGlyArgProAlaThrAl 425
QY 1112 GCCGACCATGATGCGCTCGAGCGCTGCGCGAGCGAGCGCGCGCGCGCGCTGT 1171
Db 425 AlaAlaSerAlaProAlaAlaTrpValSerAlaAlaAlaGlyArgThr-----SerSerAlaProI 444
QY 1172 G---GGGACACTTCCCGCGGTGCTGGGACATGCGCGCGAGCT-----1214
Db 444 eAlaGlySerAlaProGlySerGlyThrAlaProArgCysHisProValaArgLysAspG 464
QY 1215 -----GGGCGTGT-----TCCGCTGAGACGCGCGGTATGGAAGATGACCGCGC 1258
Db 464 yAlaGlyProAlaSerThrGlyArgSerArgArgArgTrpAlaAlaAlaProAlaArg 484
QY 1259 -----TGACGCGCG-----1267
Db 484 GAlaArgThrGlyLeuArgArgThrAlaAlaAlaCysGlyArgProProAlaProG 504
QY 1268 -----CGCGCTGCGCGCTGCGCGCGCGCGCGCG 1293
Db 504 yArgArgThrProAlaAlaAlaArgArgSerAlaArgThrAlaProArgProAlaIle 524
QY 1294 CAGTGCAGCGCGGCTACTTCCGCGAGCTGGTGTTCGACCGCGCGCGAGGTGGCGAT 1353
Db 524 AlaIleArgProSerAlaSerAlaThrGlyTrp-----ProGlyTrpAlaLeuAlaCys 542
QY 1354 ACCGCGACCTTGAACACCTTACGAGCGCGCGCGCATTCCTGTGTACGTCAAC 1413
Db 542 sProAlaAlaAlaGlyArgArgProSerProAlaProAlaProAlaProAlaProAla 562
QY 1414 GCGCGCGCGTGTGCGCAAGACAG-----CGTTACGCGCGCGACATGCGCGCGGTG 1467

```

```

Db 562 aArgProAlaAlaProAlaGlyArgProGlyArgSerProAlaThrAlaProAlaPh 582
QY 1468 CTCGACAGCGAGCGCGCTGAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1527
Db 582 eAlaAlaProAlaGlyArgProThrProAlaArgProPro-----AlaAlaAlaThrGlyArg 600
QY 1528 GCGTCCCGCGCGCTCCCAACCTGAGCGCAACCGCTACATGGCCCGCTCCCGCTGC 1587
Db 600 gProSerAlaProProAlaArgProAlaArgArgValaArgTrpProThrArgSerAla 620
QY 1588 AATA-----CGCGCCCGCGCGATATCGTGGCGCAAGAGTGTATGGCGCGCGCTG 1638
Db 620 gCysSerProAlaArgProThrProArgSer-----630
QY 1639 CGCGCGGAGCGAGCGCGCGGAAATGACCTGCAAGACTTCCGAGCGCGCGCATC 1698
Db 631 -AlaProGlyAlaAlaAlaThrGlyAla---ProThrThrGlyAlaAlaAlaArgPro 646
QY 1699 GCGGTCTCGACCTGTCCCAAGCGCGAGCTGGCGCGCGATCG 1738
Db 647 -ArgArgArgAlaAlaAlaAlaAlaProGlyLysArgArgSer 659

RESULT 9
US-09-252-991A-21920
; Sequence 21920, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rudenfeld et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMC
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21920
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21920

Alignment Scores:
Pred. No.: 4,646-16 Length: 822
Score: 431.50 Matches: 223
Percent Similarity: 35.848 Conservative: 34
Best Local Similarity: 31.108 Mismatches: 233
Query Match: 13,088 Indels: 227
DB: 16 Gaps: 42

US-10-009-782-1 (1-1758) x US-09-252-991A-21920 (1-822)
QY 45 CGATTCGACCGCTTGCAGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCG 104
Db 2 ArgArgProValaArgGlnProGlyProAlaArgArgHisProArgArgArg 18
QY 105 CCGGCGCGCGCGCGCGA-----CGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 152
Db 19 ProGlyLeuArgArgArgGlyAlaArgProGlyLeuArgArgArgArgArgAlaAlaGlnArg 38
QY 153 CATCGCGAGATCTCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 212
Db 39 GlnArgArgGlnProArgGlnAlaGlyGlnProGlyAlaAlaGlyAspHisArg 58
QY 213 CCGCGCGCGCTTATGACTCGACACCGACG-----245
Db 59 GlnProArgGlnArgLeuArgHisArgHisProArgGlnProGlyArgHisGlnProAlaPro 78
QY 246 ---GCACAACTACCTGCTAG---GCGTCGCGACATGACCGCGCAAGATCTCGAGCGCGCT 299
Db 79 GlyArgArgArgProAlaGlnProAlaGlyArgProAspLeuGln-----ProGlyArg 96

```

```

OY 300 CACGACGGTGTGTCACGGGCAATTGGGGCATCAAGCTGGC-----GCCGCT 344
DB 97 ArgProArgGlyAlaGlyArg-----GlnProGlyGlnProProGlyHisAlaPro 113
OY 345 GGC-----GCCGCGCAACCGCGCGCGCGCT 371
DB 114 GtYProGlyArgGlnProAlaAlaTyrArgTyrArgAlaGlyGlnProGlyArgProPro 133
OY 372 GGACCTGTGACGA-----AGCGGCTCTTACCGTTT 404
DB 134 GtYGlnProGlyArgGlnProArgProAlaGlnArgGlnProArgGlnArgArgArg 153
OY 405 CAGAGCGCTCGCGCACTACTGAGCGGCTGGCGGCGGCGCGCGCGCTCAAGCGCG 464
DB 154 ArgAlaGlnGlnArgGlnGlyLeuAlaAspAlaGlyHisArgAlaValaArgGlnGln 172
OY 465 CTGTATGTGGGCAATTCAAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 524
DB 173 -----ArgArgArgHisProGlyAlaValaAlaGlyAsp 183
OY 525 CACCGACGAGAAATCGCGCGCAT---GGCGGACGTGGCGGAGGAGCGATGGCGCGG 581
DB 184 SerArgArgAlaArgArgArgAlaGlnProAlaGlyProSerLeuGlyAlaGlyArgArgGln 203
OY 582 CGGCATCGGCATTCGACCGCGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCG 641
DB 204 ProHisArgHis-----ArgArgLeu-----ArgGlnPro----- 213
OY 642 GATCATCGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 701
DB 214 -----GlyTyrArgProLeuArgGlnArgProAla 223
OY 702 CGACGAGGCGA-----GCACATCGTGGCGCGCGCTGA 724
DB 224 GlnProArgArgProAlaPheProGlnProGlyArgGlyGlySerGlyArgGlnGly 243
OY 735 GGAACCTTCGCGATCGCGCGCGCGCGAGCTGGA---CGTGGCGGTGTATCTCGACCA 791
DB 244 ArgArgArgAlaHisArgLeuGlnProGlyArgArgAlaGlyGlnProLeuArgProAla 263
OY 792 GGCAT-----GGCGGCGCGCAATTTCGGCGCGCGCGCGCGCGCGCGCGCT 839
DB 264 GtYArgArgGlyArgAlaAlaArgAlaArgArg----- 275
OY 840 GATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 878
DB 276 AspArgGlnGlnArgArgGlnProAlaArgProArgProGlnArgGlnHisAlaAlaGly 295
OY 879 CGCGTATCCCTAGCTGGCGCGCGCTCCACAGCT-----CAAGCGAGCGCGCGCT 929
DB 296 ArgTyrArgProGlnGlnArgLeuArgArgAlaGlyLysArgGlnProGlyProGlyPro 315
OY 930 GCT-----GGCGGAGCGCACCATCATCATCGTGGCAAGCGCTT 968
DB 316 AlaThrGlyGlnProGlyGlnArgArgArgAlaHisProProHis----- 330
OY 969 CCGCGACGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1007
DB 331 ---ArgGlnTrpHisLeuArgProGlyPheArgAlaGlyAspProArgArgArgThr 349
OY 1008 -----CGAGCGCGCAATTCAGAGAGAGAGAGAGCGCGCG 1040
DB 350 AspHisGlnTrpProAlaGlyHisProArgGlnArgGlnArgGlnGlnGlnGlnGln 369
OY 1041 CGAGCTGCA-----GGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1094
DB 370 SerArgThrProGlnProGlyHisArgHisLeuProProAspGlyArgGlyAlaAla 389
OY 1095 GCGCATCTCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1145
DB 390 GtYGlyAlaValaLeuHisTrpProArgArgArgLeuGlnGlnArgArgProAlaGlyGln 409
OY 1145 CAGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1205

```

```

DB 410 ArgArgGlnLeuAlaProArgProGlyArgGlyPro-----Trp 423
OY 1206 GCGCGCGCTGGCGCTGTTCGCGCTGGAGAGCGCGGATGAGATGACCGCGCTGACCG 1265
DB 424 GlnArgProArgHisGlnPro-----ArgArgLeuArgProGlyArg 437
OY 1266 CGCGCGCTTCGCGCT-----GCCGCG 1286
DB 438 ArgGlnProArgProArgGlnArgGlnArgGlnProArgArgArgGlnGlnHisAlaArg 457
OY 1287 GCGCGCGGACGTGACAGCGCGGGA-----CTTCGCGCGCGCGCTGTGT 1331
DB 458 ArgArgGlnProAlaGlyGlnProTrpAlaAspHisArgArgArgArgProArgGlyGln 477
OY 1332 CGACCGCGCGACGCTGGCGGAGAACCGCGCGCGCGCGCGCGCGCGCGCGCG 1391
DB 478 Arg-----ArgGlnSerGlnGlnLeuArgHisPro-----GtYArgArgArg 491
OY 1392 CATCCATTCCGCTTACGTACAGCGCGCGCGCGCTGGCGAGCGAGCGCT----- 1442
DB 492 GlnProAlaAlaGlyArgAlaArgProAlaGlnArgAlaArgAlaValaGlnTrpArg 511
OY 1443 ---CACGCGCGACGA-----TGC----- 1457
DB 512 ArgHisAspProAlaArgArgArgHisGlnProLeuArgGlyCysValaGlnProArg 531
OY 1458 -----CGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1505
DB 532 GlnAlaGlyTyrArgProArgArgArgCysArgGlnProCysArgGlnProAlaGlnProPhe 551
OY 1506 ACAATCGCGCGGTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1565
DB 552 ArgGlyAspArgGlnArgGlnGlyLeuGlnProAlaCysGlnProAspArgGly 569
OY 1566 CATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1619
DB 570 -----ProSerArgArgAlaGlyLysGlnValaGlyProVal---HisArgGlyAspGly 586
OY 1620 AGTATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1658
DB 587 AlaAsp---ArgLeuHisArgArgArgGlnArgGlyArgLeuGlnArgGlnAlaGlnArg 605
OY 1659 -----GAAATGACCTCGAAGACCTGTCGCGCGCGCGCGCGCGCGCGCGCG 1712
DB 606 HisLeuAspHisTrpProAlaArgGlnAspArgGlyHisArgGlnGlnArgHisGlyAla 625
OY 1713 GTCCAGCG-----CGAGCTGGCGCGCGCGCGCGCT 1742
DB 626 ThrAlaGlyArgArgArgLeuArgHisArgArgArgHisProGlnGlnPro 642

```

```

RESULT 10
US-09-252-991A-32204
; Sequence 32204, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: MATE J. RUBENFIELD et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONA
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32204
; LENGTH: 651
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32204
Alignment Scores:

```



Pred. No.: 5, 11e-16  
 Score: 430.50  
 Percent Similarity: 34.57%  
 Best Local Similarity: 29.59%  
 Query Match: 13.05%  
 Matches: 208  
 Conservative: 35  
 Mismatches: 247  
 Indels: 213  
 Gaps: 38

US-10-009-782-1 (1-1758) x US-09-252-991A-32204 (1-651)

```

QY 54 GCGCTTCGACCTGCTGCTGCGGCGGACACCTCATGACGAGCAACACCCGGGGG 113
DB 5 AAGAGLNGLProlglaArgProvalProhArgValaAspHisArgGlyArg 24
QY 114 GCGCGGCGACCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 173
DB 25 ProlglaArgGlnaArgGlyArgProGlnaArgGlnaArgGlnaArgGlnaArg 43
QY 174 GCGCGGCGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 233
DB 44 ArgArgSer-----ProGlyValSerGlnProValaAlaArgLeu 56
QY 234 GCAACCCGACGAGACACTGCTGCT-----CAGGCGTCCGACATGAGCGC---CAA 284
DB 57 CysAlaProGlnaArgAlaArgProArgProhHisArgProhArgAspAlaLeu 76
QY 285 GATTCGACGAGGCGTCAACAGGTGTACGCGGCA-----TTGCGGCAAT 329
DB 77 ProvalProhArgArgHisArgProhHisArgArgSerArgProAlaArgHis 96
QY 330 CAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 362
DB 97 ArgProGly-----GlyGlyArgArgProAlaAlaGlyArgTyrGlnGlyValGly 114
QY 363 GCGCGGCGTGCACCTGCTGCGGAGAAAGCGGCGGCTTACCGTTGCGGCGGCGGCGG 422
DB 115 ArgHisArgGlyProGlyArg-----ArgArgGln 125
QY 423 CCGTGCAGCGGCTGCGG-----GGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 476
DB 126 ProGlyAsnArgAlaGlnGlyGlnHisGlyGlyAspArgValaArgArgProGlyAla 145
QY 477 CCATTCACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 518
DB 146 -----GlyAlaArgArgGlyArgGlyProAlaGlnaArgProhArgProArgGln 162
QY 519 -----CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 560
DB 163 ProProAlaProhHisArgArgProhArgArgGlnaArgGlnaArgAlaSerProAla 182
QY 561 CGAGGAAC-----569
DB 183 ArgHisAlaAlaHisArgArgProGlyAlaArgProhArgProhArgProhArgGly 202
QY 570 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 629
DB 203 AsnAlaGlnaArgArgGlnaArgGlnaArgGlnaArgGlnaArgGlnaArgGlna 221
QY 630 CAGCAGGAGATCATGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 689
DB 222 SerAlaGlnaArgGlyArgGlyProArgGlyArgProhArgProhArgHisArgHis 241
QY 690 CAGCCGACATGCGGCA-----CGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 731
DB 242 AlaProhHisArgArgArgGlnaArgGlnaArgGlnaArgGlnaArgGlnaArgGlna 261
QY 732 -----GAGGAACCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 785
DB 262 ValArgGlyGlyArgLeuProhHis-----ArgAlaSerGlyGlnaArgAla 276
QY 786 CAGCAGGCTCATGG-----CCAGCCGCAATTCGCGGCGGCGGCGGCGGCGGCGG 833
DB 277 ProAlaAlaHisGlyGlyHisAlaAspHisArgGlnaArgProvalaLeuArgArgTle 296
  
```

```

QY 834 GCGGCTGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 869
DB 297 ArgArgArgLeuArgArgArgArgArgArgArgArgArgArgArgArgArgArg 316
QY 870 -----CTGCGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 920
DB 317 LeuArgLeuProhHisArgArgArgArgArgArgArgArgArgArgArgArgArg 336
QY 921 CCGCGTGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 953
DB 337 GlnaArgLeuProhArgGlnaArgGlnaArgGlnaArgGlnaArgGlnaArgGlna 356
QY 954 CTGGTCAAGCCGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1013
DB 357 LeuGlnaArgLeuProhArgArgArgArgArgArgArgArgArgArgArgArgArg 373
QY 1014 CCGCAAAATCAAGTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1058
DB 374 ArgGlyAspHisArgArgArgArgArgArgArgArgArgArgArgArgArgArg 393
QY 1059 GCGCATCTACTCATGATGAGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1118
DB 394 AlaGlnaLeuArgGlnaArgGlnaArgGlnaArgGlnaArgGlnaArgGlnaArg 409
QY 1119 CATGATCGGCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1169
DB 410 ---AspArgArgAlaGlnaArgProvalaGlyArgAspArgArgAlaGlySerGlnaArg 428
QY 1170 GTGGGCGACCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1226
DB 429 ValArgProProhArgGlyAsnProCysLeuGlnaArgGlnaArgGlnaArgGlna 446
QY 1227 GCTGAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1286
DB 447 -----GlnaArgPro-----GlyAlaAspProGlyGln 456
QY 1287 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1346
DB 457 ValArgHisArgLeuAlaHisArgArgProhArgArgGlnaArgGlnaArgGlna 476
QY 1347 GCGCGATACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1385
DB 477 TyrArgArgGlnaHisProProGlnaArgArgArgArgArgArgArgArgArg 496
QY 1386 -----CGCGG-----CATTCATTCGCT 1403
DB 497 GlnaArgProGlyArgArgGlyLeuProGlyArgArgGlyAlaHisArgProhHis 516
QY 1404 GTACGTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1448
DB 517 ArgHisGlnaArgGlnaArgGlnaArgGlnaArgGlnaArgGlnaArgGlnaArg 536
QY 1449 CAGCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1508
DB 537 ProGlnaArgGlnaArgGlnaArgGlnaArgGlnaArgGlnaArgGlnaArgGlna 555
QY 1509 ATCCGCGG-----GAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1553
DB 556 ValArgArgAlaAlaLeuGlnaArgGlnaArgGlnaArgGlnaArgGlnaArgGlna 575
QY 1554 CGGAAACGCTACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1577
DB 576 AlaTyrProArgAlaAspProAlaValaArgGlyArgGlyProAlaGlyLeuLeuArgPro 595
QY 1578 -----CTCGGCTCGCAATAC 1592
DB 596 GlyProValaGlnaHisArgLeuProhHisArgGlnaArgGlnaArgGlnaArgGlna 615
QY 1593 GCGCGC-----ACCGGATATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1628
DB 616 GlyProTyrProAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 635
QY 1629 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1688
  
```



```

OY 1242 ATGAGATGACGGCTGACCGCGCGCTTCGGCTGCGCGCGCGCGCGAGCTGCA 1301
Db 521 -----ProserArgArg-----ArgProAlaPheAlaValGluProGln 533
OY 1302 GCGCGGCTACTGCGCGACCTGGT---GATGTCGACCGCGCGCGCGCGAGCTGCA 1358
Db 534 ArgArgLeuAlaGlnArgProGlyAlaGlyValArgProAlaAlaLeuProArgArg 553
OY 1359 CACCTTCGAAACACCTACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1418
Db 554 ThrProLeuArgProProAlaProAlaProGlnProValSerProAlaArgArgArg 573
OY 1419 GCGCGTCTGCGAAGACAGCGCGCTCAC-----CGCGCAGCA 1454
Db 574 ProGlyGlnProGlyAlaArgCysHisProSerLeuGlyValAlaProArgGlnPro 593
OY 1455 TCGCGCGCGCGT-----GCTGCGACGACGCGCGCGCGCGCGCGCGCGCGCGCTTACA 1508
Db 594 ILeuAlaProArgMetGlyLysArgGlnHisArgSerArgArgProAsp----- 611
OY 1509 ATCCGCGCGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1568
Db 612 ---AlaArgGlyArgGlyLysArgProCysProProGlyArgPheGluArgAlaAsp----- 628
OY 1569 GCGCGCTCCCTCCGCTCCCAATAC-----GCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1622
Db 629 -----GluArgGlnTyrArgAlaGlyProGlyArgAlaArgProAspGlyArg 644
OY 1623 GATGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1682
Db 645 SerArgArgHisProGly---GlyAlaGlyLysArgPheGlnHisProArg 663
OY 1683 GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1739
Db 664 AlaProArgArgGlySerGlyLysAlaValLeuGlnArgGlyArgGlyArgGlyArg 683

RESULT 12
US-09-252-991A-25557
; Sequence 25557, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ. ID NOS: 33142
; SEQ. ID NO 25557
; LENGTH: 1149
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25557

```

```

Alignment Scores:
Pred. No.: 1,21e-15 Length: 1149
Score: 424.50 Matches: 196
Percent Similarity: 34.75% Conservative: 25
Best Local Similarity: 30.82% Mismatches: 224
Query Match: 12.87% Indels: 191
DB: 16 Gaps: 33

```

US-10-009-782-1 (1-1758) x US-09-252-991A-25557 (1-1149)

```

OY 15 GCGGAGAGAGAGATTTCATGTCCTCA---ATCCGATCCGACGCGCTTCGACCTGCTCT 71
Db 152 ArgAlaHisArgGlnProSerAlaProValAlaArgArgProAlaAlaAlaGlyValAsp 171
OY 72 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 131
Db 72 GlnProHisArgHisProLeu-----ArgCysGlnHisProPro 488

```

```

Db 172 ArgProArgAlaAspGluArgArgProTyrHisPro---ArgArgArgGlnHisArgArg 190
OY 132 GCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 191
Db 191 ProArgGlnProGlnArgGlyLysSerHisGlyAlaAlaArgArgAlaGlyLysGlnProGly 210
OY 192 CGACGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 251
Db 211 ProArgGlyLysProAsp-----His-----ProArgProArgArgArg 223
OY 252 CTACGTCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 311
Db 224 ArgProArgGlnAla-----HisHisArgGly 232
OY 312 -----CACGCGCAATTGCGCGCATACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 359
Db 233 AlaArgArgArgArgArgArgGlnArgGlnArgGlnArgGlnArgGlnArgGlnArgGln 252
OY 360 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 416
Db 253 ArgArgArgAlaProProArgAlaGlyArgArgSerGlnProAlaProArgArgGlyGln 272
OY 417 CGACTACCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 473
Db 273 ArgGlnLeuGlyLysLeuAlaArgArgGlnArgGlnArgGlnArgGlnArgGlnArgGln 292
OY 474 GGG-----CAATTCAAGCT-----GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 512
Db 293 AspValAspGlnSerValProHisArgAlaAspProAlaArgHisHisArgArgArg 312
OY 513 GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 572
Db 313 LeuGlyLysGlnHisAlaArgArgArgArgGlyGlnGlnAlaProGlyAspGlyAlaAsp 332
OY 573 GCGCGAG---GCGCGCGCATTCGCGCGCGCGCGCTTA-----CGCGCGCGCGCG 623
Db 333 GlyArgValArgLeuGlnHisLeuSerGlnArgLeuAlaGlnProAlaArgAla 352
OY 624 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 683
Db 353 AspGlyHisArgGlnGln-----358
OY 684 CTACGCCACCCACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 743
Db 359 -----ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 372
OY 744 CCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 803
Db 373 GlnHisAlaGlyLysArgArgArgAlaGlyAlaLeu-----ArgGlnHisArgLeu 390
OY 804 GCGCAATTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 863
Db 391 ProArgLeuArg-----ArgArgGlnGlnHisArgLeuPro 402
OY 864 GGA-----GTCGTCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 917
Db 403 GlyAsnProGlnLeuAlaGlyLys-----ArgGlyGlnLeuLeuHis----- 416
OY 918 GACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 968
Db 417 -----ArgAlaArgArgArgArgArgHisHisGlyArgGlyAspArgLeuGlnGlyAla 434
OY 969 -----CCCGGAATGACGCGCGCG 986
Db 435 GlnGluAlaValArgGlnArgGlnProAspArgProLeuHisArgHisArgGlnArgAla 454
OY 987 CGACTGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1046
Db 455 ValProGlnHisArgArgAlaArgArgGlnArgGlnLeuGlnLeuArgGlnGlnCysArg 474
OY 1047 GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1106
Db 475 GlnProHisArgHisProLeu-----ArgCysGlnHisProPro 488

```

```

OY 1107 GTTCGGCCCGACCATGATGCGCTCCGACGCGCTGCGCCGACGACGACGCGCCGATCCGGC 1166
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 489 VALATGHSALAGLInProAlaArgValAlaArgAspHisArgAlaArgAlaArgProAlaArgAla 508
OY 1167 CCT-----GTGGGCGACCTTCGCGCGGCTGGGGGCTGAGGCGACTATGCGCGGACGTGGGCGCT 1220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 509 ProGlyArgThrAlaArgHisArgProValAlaAlaAla-----ProAlaPro 523
OY 1221 GTTCCCGCT-----GAGACGCGCGGTATGGAAGATGACCGCGCTGACCGCGCGCG 1271
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 524 ArgProAlaArgLeuAlaArgAlaAspGlnAlaArgGlyAspAsp-----537
OY 1272 CTTCGGCGCTGGCGCGCGCGCG-----GCAAGCTCAGCGCGCGGTACTTCCG 1316
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 538 -----ProGlyArgGlyGlyAlaAspProGlnHisProValAlaAlaAspAlaArgLeuAspArg 555
OY 1317 CGA-----CTGTGTGTGTTCGACCCGCGCCACGAGTGGCGCGCATCCGCGCTTCGACGA 1370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 556 ArgAspLeuProAlaGlyArgAlaArgAspArgAlaAspGlnHisAlaHisAspAla 575
OY 1371 CCTTACCGACGCGCGCGCGCGCATTCCTGCTAGCTACGACGCGCGCGCGCTGGCA 1430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 576 ArgThrHisProArgAspArgHisPro-----584
OY 1431 AGAGACGGGCTCACCGCGCACGATGCGCGCGCGCTGCTGCACGACGCGCGCGCTGAGC 1490
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 585 -----HisGlyHisThrArg 589
OY 1491 CCGGCGCGACGCTTACATCCGCGCGGTGAACGCGCGCGCGCGCGCGCTCCGACACCGT 1550
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 590 ProPro-----GlyArgTyrProProProValPro 599
OY 1551 GAGCGCAACCGCTACATGCGCGCGCGCTCCGCTCGCAATGACGCGCGCGCGCATTCGT 1610
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 600 HisArg-----Gly 602
OY 1611 GGGCAGAGAAATGATGGCGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCG-----GAA 1661
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 603 GlyAspAlaLeuGlyGlyArgArgProGlyArgAspArgProGlyProValHisArgArg 622
OY 1662 AATGACCGCTGACGACCTGTCCGACGCGCGCATGCGCGCGCTGCGACCGCTGTC-----1715
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 623 ArgAlaAlaAlaArgProGlyArgGly-----GlyLeuPheProValGlyHis 638
OY 1716 -----CAAGCGCGAGCTGGCGCGCATGCGCGCTGACGCTACGA 1751
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 639 ArgArgArgLeuGlnLeuArgAlaGly---HisArgProGlyValArg 653

```

```

RESULT 13
US-09-252-991A-29274
; Sequence 29274, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29274
; LENGTH: 801
; TYPE: PRT.
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29274

```

Alignment Scores: 1.4e-15 Length: 801  
Pred. No.: 423.00 Matches: 222

```

Percent Similarity: 35.15% Conservative: 29
Best Local Similarity: 31.09% Mismatches: 253
Query Match: 12.82% Indels: 210
DB: 16 Gaps: 40

US-10-009-782-1 (1-1758) x US-09-252-991A-29274 (1-801)
OY 15 CCGGGAAGAGAGATTCATGCTCCCAATCCAGATCCAGCCCTTGAGCTGCTCCG 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 116 ArgGlyHisArgGln-----ProTyrProAlaArgAlaArgProAlaArg 130
OY 75 GGGCGCGACCTTCATGAGCGGCGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 131 AspArgHisArgArgArgArg-----ProGlySerGlyAspArgAlaArgPro 147
OY 129 -----CGTGGCGGGGAGCG-----143
DB 148 ArgLeuLeuAspAlaProAlaGlyHisArgAlaArgAlaArgAlaArgProAla 167
OY 144 ---CATGCGCGCGCATGCGCGCGATCTGGAACCGCGCG-----GCACACCG 188
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 168 GlyHisHisArgHisArgArgGlyAlaGlyProAspArgValProAlaGlnAlaGly 187
OY 189 GGTGACGCTGTGCGCGCGCTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 188 GlyArgArgLeu---ProGlyAlaLeuArgArgArgArgLeuGlnProAspProArgArg 206
OY 249 CAAGTACCTGCTCAGCGCTGCGACATGACGCGCAAGATCTGCGACGCGCGCGCGCGCGT 308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 207 -----ProArgAspHisLeuGlnHisGlySer-----GlyLeuArgHis 220
OY 309 GGTACGCGCGCAATGGCGCATGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 221 ---ArgGlyAspValArgHisArgPro-----AlaAspHisArgLeuProHis 237
OY 363 -----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 238 ArgProArgArgArgThrGlyArgProGlyGlyLeuGlyAlaHisArg 254
OY 402 TTTCGACGCTTCGCGCGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 255 ---ArgThrLeuGlyArgGlnProGlyArgArgValArgAlaGlyThrGlyValArg 273
OY 462 CGCGCTG-----TATGGTGGCGCATTCACGCTGCGCGCGCGCGCGCGT---CAT 503
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 274 ProValGlnArgGlyAlaGlnHisGlyArgProValGlnSerAlaArgGlnGlyHis 293
OY 504 GCGGACCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 563
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 294 GlnArgThrGlyGlyGlyArgHisArgArgGlnProArg-----AlaGlyProArgArg 311
OY 564 GGAAGCGCATGGCGCGCGCGCGCGCGCATTCGACCGCGCGCGCGCGCGCGCGCGCGCG 623
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 312 GlySerArgGlyProAspAlaGlyArgArgGlyAsp-----323
OY 624 CCGCGCGACCGAGAGATCATGAGGTGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAT 683
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 324 -----HisArgArgHisArgHisGlnHisGlnGlnProAlaGlnArgAsp 340
OY 684 CTAAGCGACCCCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 743
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 341 ArgArgArgProAlaGlyAlaGlnArgArg-----350
OY 744 CCGCATGCGCGCGCGAGCTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 791
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 351 ProProArgProGlyProGlnAlaMetGlyGlyAspLeuAlaGlyThrArgLeuGlnGly 370
OY 792 GGTGATGG---CAAGCCCAATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 836
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 371 GlyHisArgGlyLeuProAlaArgSerArgProAlaAlaAlaProGlySerProArgLeuArg 390
OY 837 -----GCTGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

OY      1683 GGAGGCGCAGCGCATCCGGCTGCACCCCTGCCAAGGCCGA 1724
        |||   ...    |||   |||   |||   |||
Db       750 ALaAlaArgSerAlaGlyAsrLeuProPrgIugInaArg 763

RESULT 14
US-09-252-991A-26099
; Sequence 26099, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ. ID NOS: 33142
; SEQ. ID NO 26099
; LENGTH: 863
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26099

Alignment Scores:
Pred. No.:          1,72e-15           Length:         863
Score:              421.50             Matches:         232
Percent Similarity: 32.06%             Conservative:    29
Best Local Similarity: 28.50%          Mismatches:     264
Query Match:        12,788             Gaps:           289
                                      Indels:         41

US-10-009-782-1 (1-1758) x US-09-252-991A-26099 (1-863)

OY      6 CCAATTGATCGSCGAGAGAGAT-----TTCAATGTCCCAATCCGATTCGACGCC 56
        |||   |||   |||   |||   :|||::|||
Db      48 ProSeclnArgTgltArgGlyAsrProrPrgAluGluArgLeuArgTgluArgTg 67
OY      57 CTTCGACTGTCTGTCTCGCGGCGCACCT-----CATCGAGCGCAGAACACCCGCG 110
        ::|||::|||   |||   |||   :|||::|||
Db      68 lleaPrgocysLeuArgSerAlaGlyAsrCyshPheNlaArg-----Prglly 83
OY      111 GCGGCGCGCGGCACT-----GGCGGTGCGCGCGCACCGCATCGCGGCAATCGGAT 164
        |||   |||   |||::|||
Db      84 CySgLySerHisrProArPrgIugInaLarProValGluaSargHnAlaArg 103
OY      165 GTCSGACGC-----CGCGGCGCACACCGGGTGCAGCTGTGCGGCTGGT 209
        |||   |||   |||::|||
Db      104 ArgGlyAlaValHisrProArPrgIugLeuArgTgltArgTgHnGlyArgTg---AsrPrgAla 122
OY      210 GGTGCGCGCGGCTTCATCGACT-----GCACACCCGACGACGACATA 254
        |||   |||   |||::|||
Db      123 AlaAlaGlyAluArgTgHnIsrArgLeuPrgroThnAsrSargIneIugInaArgAluArgTg 142
OY      255 CGTGCTACGGC-----TCGCGACATGACGCCCAAGAATCTCGCAGGCGCTCAC 302
        |||   |||   |||::|||
Db      143 ProArgHnIsrArProGlyIugLyIgtHnArgProThnArgAlaGlyProValAlaHnIsrGthr 162
OY      303 CACGCGTGCTCACGGGCAATTCGGCGCATCAG-----CCTGGCGCGCGCT----- 344
        |||   |||::|||::|||   ::|||
Db      163 LeuGlySerAlaGlyArgLeuArgTglnArgTgTrnAsrProGlyIugIuArgAlaAla 182
OY      345 -----GGCGCACGCGACACCCGCGCGCGCGCGCTGACCTGGTGGGA 383
        |||   |||   |||::|||
Db      183 ProAlaProCysGlyAsrPrgoGlyValAlaArgPrgroArProGrgInPrgAlaLeuPrgAlu 202
OY      384 CGAAGCGGCGCTCTTA-----CGCTTTCGAGCGCTTGGC-----CGACTAACT 425
        |||   |||::|||::|||   |||
Db      203 ArgThrAlaArgLeuGlyAlaAlaProProGrgIugIuHnIsrGlyIuSerSerLeuPro 222
OY      426 GGA-----GCGGTTCGG-----GGCGACGCGCGCGCGCGCTTAAGCGCGCGCGCG 457

```

Db 548 ThgIyArGgIyHnIsPrOaYgArGgIaYgPrOoGlyArGdYgInPrhArGhInIa 567  
 QY 1185 -----GCGGTCGTGGGCACTATGGCGCGA 1211  
 Db 568 GInGIyAlaArGlyArGArGhInArGArGgIyAlaInIaArGArGgIyArGThrGlyArG 1111  
 QY 1212 CCGTGGCGCTGTCCCGGTGGAGCGGCGTATGGAAAGT----- 1250  
 Db 588 GInGIyPrOThrAlaAlaInIaGlyAlaGlyIyAlaAmnIaArGgIyThrPrOAlaInGIyArG 1111  
 QY 1251 -----GACCGGCTGACGCGCGCGCTTGGCGTGGCG 1283  
 Db 608 SerArGArGArGgIyPrOArGArGPrOoGlyArGPrOThrAlaInIaGlyAlaInPrOAla 1111  
 QY 1284 CGGCGCGGGGAGGTGGAGCGCGGCTACTTCGCGCACTGTGTTCGACCGCGCAC 1343  
 Db 628 GLyAlaInIaArGgIyGlySerArGArGLeuArGhIsPrOAla-----AlaHIs 1111  
 QY 1344 GGTGGCGGATACGCG----- 1358  
 Db 644 PrOoGInArGgIyInArGInGIyAlaInGIyArPrAlaHIsValThrAlaArGInArPrOAl 1111  
 QY 1359 ---CACTTGGAAACACC-----TACGAGCGCGCGC 1388  
 Db 664 HIsHIsArGArGgIyInPrOArGhIsArGPrOoGlyArSPrPrOAlaArGArGAlaArGTPr 1111  
 QY 1389 CGGATTCATTCGCTGTACCTCAACGCGCGCGGCTGTGCGAAGAGCGCTTCACCGG 1448  
 Db 684 ArGgIyPrOoGlyArSPrArGgIyGInGIyArGArGAlaAlaSerGIyAlaGlyArGhInIs 1111  
 QY 1449 CCGCATGCCGCGCGCGCTGTCCACGACGCGCGCTGAGCGCGCGCGCGCGCTTACA 1508  
 Db 704 LeuGIaArGArGArGArGgIyArG-----GlyArGArGArGPrOoGlyPrOAlaAlaAla 1111  
 QY 1509 ATCCGCGGTGAACGCGCGCGCGCTGCCCGCCCTCCACCGCTGGAGCGCAACCGCTAACAT 1568  
 Db 722 ThrArGArGAlaArG-----ArgAlaInIaArGgIyGlyGInIs 1111  
 QY 1569 GGGCCCTCCCTCGCTCG-----CAATACGGCGCCACCGATAT----- 1607  
 Db 735 GlyPrOArGArGArGThrLeuArGArGArGArGAlaGlyInGIyAlaInArGArGgIyArG 1111  
 QY 1608 ---CGTGGCAAGAATGATGGCGCGCGCTGGCGCGCGCGAGCGCAAGCGCCGGAAT 1664  
 Db 755 PrOPrOoGlyArGgIyGlyAlaInIaGlyAlaInIaLeuArPrOArPAlPrOArPAln 1111  
 QY 1665 -----GACCTCGCAAGACTGTGCGAGCGCGCGATCGGCC 1703  
 Db 775 HIsPrOArGlyAlaGlyLeuGIyPrOoGlyArGPrOAlaLeu--PrOAlaInGIyArG 1111  
 QY 1704 CTCGACCTGTGTCCAGCGCGGAGCTGGGCCATCGGCC 1741  
 Db 794 InArGPrOhIsPrOoGInPrOAlaAlaAlaAsPInPrO 806  
 RESULT 15  
 US-09-252-991A-16789  
 ; Sequence 16789, Application US/09252991A  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 PRIOR FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 16789  
 LENGTH: 774  
 TYPE: PRY  
 ORGANISM: Pseudomonas aeruginosa



US-09-252-991A-16789

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
4.81e-15	413.50	33.088	28.708	12.538	774	223	34	255	265	41

US-10-009-782-1 (1-1758) x US-09-252-991A-16789 (1-774)

```

QY 18 GGAAGAGAGATTCATGCTCCATTCGATTCGACCCCTTCGACCTGCTGCTGCGGG 77
Db 58 GATAGGlnAaP-----LeuProleuLeuInProProGlnSerArg 71
QY 78 CGGACCCCTCATCGACGCGACGA-----CACCGCGCGCGCGCGCGCGA 122
Db 72 ProAspProGlnArgSerArgInAlaAlaAspValAlaGlnGlyProAlaGlyLysPro 91
QY 123 CCGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
Db 92 ProAlaLeuGlnGlyArgGlnInAlaArgInAlaArgInAlaArgInAlaArgInAla 111
QY 183 -----CACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 218
Db 112 AlaGlnGlnAlaProProArgProValProProGlnAlaSerProAla 131
QY 219 CGGCTTCATCGACGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 278
Db 132 GlnLeuInAlaArgGlnAlaArgGlnGlyLys----- 141
QY 279 GCCCAAGATTCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 338
Db 142 -----ProGlnArgInAlaAlaArgGlnGlnGlnGlnGlnGlnGlnGlnGln 156
QY 339 GCCCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 398
Db 157 -----AlaGlnArgInProAlaValAlaGlnGlnGlnGlnGlnGlnGlnGln 173
QY 399 CCGCTTCGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 443
Db 174 AspGlnArgGlnValAlaArgGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 193
QY 444 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
Db 194 AlaGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 212
QY 501 CATGCGGAGCTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 557
Db 213 GlnArgGlnGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGln 232
QY 558 -----GCGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 602
Db 233 HisGlnLeuAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 622
QY 603 CGCGCTTCACCC-----GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 644
Db 253 HisArgLeuProInHisHisHisArgGlnAlaArgProAlaArgGlnAlaArgGln 272
QY 645 CATGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 689
Db 273 HisArgGln-----GlySerArgGlnAlaArgProAlaGlnGlnGlnGlnGln 286
QY 690 ---CACCGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 746
Db 287 AlaAspProArgGlnAlaArgGlnAlaArgGlnAlaArgGlnAlaArgGlnAlaArg 306
QY 747 CATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 782
Db 307 HisArgProGlnAlaAspArgInHisHisHisArgAlaAlaGlnGlnGlnGlnGln 326
QY 783 -----GCACCGACAAAGT----- 794

```

```

Db 327 ArgArgValLeuArgArgArgProGlnArgProAlaProGlnGlnGlnGlnGln 346
QY 795 ---CARGG----- 800
Db 347 GlnHisGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 366
QY 801 -----CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 848
Db 367 ArgLeuProAlaProValArgInHisHisHisGlnGlnGlnGlnGlnGlnGlnGln 386
QY 849 -----GCCATGCGCGCG 860
Db 387 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 406
QY 861 CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 920
Db 407 ProGlnHis-----GlyArgSerArgGlnGlnGlnGlnGlnGlnGlnGlnGln 942
QY 921 CCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 974
Db 425 ProArg-----GlyProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 442
QY 975 -----ACTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1010
Db 443 ThrAlaSerLeuGlnHisArgArgArgProProAlaGlnGlnGlnGlnGlnGlnGln 462
QY 1011 ---CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1037
Db 463 GlnArgArgGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1082
QY 1038 -----GCCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1088
Db 483 ProGlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 498
QY 1089 CGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1148
Db 499 -----HisLeuGlnProGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 513
QY 1149 -----CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1196
Db 514 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 533
QY 1197 GCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1286
Db 534 -----LeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 545
QY 1257 CCGTAC-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1295
Db 546 ProAspProLeuProArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 565
QY 1296 -----GCTGCGCGCGCG 1307
Db 566 HisArgGlnLeuArgProAlaAlaGlnAlaAspArgGlnSerHisProAlaGlnArgPro 585
QY 1308 -----GACTCGCGCGCGCG 1322
Db 586 AspArgArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 605
QY 1323 -----GGTGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
Db 606 GlnAspGlnLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 639
QY 1350 CGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1409
Db 624 -----GlnArgAlaAspGlnProGlnArgGlnGlnGlnGlnGlnGlnGln 637
QY 1410 CAA-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1466
Db 638 GlnGlnArgProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 657
QY 1467 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1511

```

```
Db 658 GlyAspProGluGlyArgHisArgAspValProGlnGlyValArgArgGlyLeuArgArg 677
QY 1512 CGCGGTGAAACGGGGCGGCGGTGCCGCCCTCC-----AACCTGGAGGC 1556
    |||||
Db 678 ArgArgGluMetAlaGlyAspProGlyAlaAlaValGlyHisLeuArgValAlaGlyArg 697
QY 1557 AAACCGCTACATGAGCCCTCCCTC-----CGCTCGCAATACGGCCC 1598
    |||||
Db 698 LeuHisLeuHisProAlaSerThrLeuLeuArgThrHisArgArgGlyAlaAlaGlyHis 717
QY 1599 ACCCGATATCGT-----GGGCAAGAAAGTATGGCGC----- 1631
    ||| |||
Db 718 ArgArgArgArgAlaGlyAlaGlyAlaGlyAlaGlyArgLeuGlyAspHisArgPro 737
QY 1632 ---GCCCTCGCGCGCGAGCGCAAGCGCGGAAATGACCTGCAGACCTGTGCGCAGGC 1688
    ||| |||||
Db 738 HisLeuProCysArgGlnHisGlnGly-----ArgGlnProGlyArgProLeuProAla 755
QY 1689 CAGCGGCATCGCGCTCGACCTGTCCAAAGCGCGAGCTGGGCCAGATCGC 1739
    ::| |||||
Db 756 ArgAlaArgArgGlyThrGluGlyLeuGlnLeuLeuArgLeuProSerArg 772
```

Search completed: May 11, 2003, 12:25:17  
Job time : 262 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: May 11, 2003, 12:16:00 ; Search time 55.5 Seconds

(without alignments)  
5829.953 Million cell updates/sec

Title: US-10-009-782-1

Sequence: 1 gattacactgacgacgcga.....ccctgacgacgacgaacct 1758

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 349150 segs, 92025710 residues

Total number of hits satisfying chosen parameters: 698300

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+2p.model -DEV=alp  
-CGN2\_1/USPTO-SPOOL/US1000782/runat\_07052003\_122319\_23259/app\_query.fasta.1.1927  
-DB=PubMed\_Applications\_AA -QMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-DOOPL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=pcr -THR.MAX=100  
-THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcr -NOR=ext -HEARSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USPR=US10009782 -CGCN\_1.1.12 -etunal\_07052003\_122319\_23259  
-NCP=6 -ICRP=3 -NO.XLPEX -NO.MAMP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications\_AA:

1: /cgn2\_6/ptodata/2/pubpa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpa/PCIT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpa/PCITUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	612	18.6	558	10 US-09-770-517C-2
2	310	9.4	558	9 US-09-975-719-277
3	296.5	9.0	1367	10 US-09-801-368-108
4	289.5	8.8	5179	9 US-10-025-380-1068

Result	Score	Match Length	DB ID	Description
5	289.5	8.8	5179	10 US-09-922-217-1068
6	289.5	8.8	5179	10 US-09-833-263-1068
7	285.5	8.7	941	12 US-10-124-557-14
8	285.5	8.7	1022	12 US-10-124-557-84
9	285.5	8.7	1038	12 US-10-124-557-74
10	285.5	8.7	1049	12 US-10-124-557-58
11	285.5	8.7	1140	12 US-10-124-557-104
12	285.5	8.7	1270	12 US-10-124-557-142
13	285.5	8.7	1313	12 US-10-124-557-44
14	285.5	8.7	1361	12 US-10-124-557-48
15	285.5	8.7	1404	12 US-10-124-557-52
16	285.5	8.7	1404	12 US-10-124-557-62
17	285.5	8.7	1464	12 US-10-060-036-159
18	285.5	8.7	1464	12 US-10-060-036-159
19	285.5	8.7	1464	12 US-10-060-036-159
20	285.5	8.7	1464	12 US-10-060-036-159
21	285.5	8.7	1464	12 US-10-060-036-159
22	285.5	8.7	1464	12 US-10-060-036-159
23	285.5	8.7	1464	12 US-10-060-036-159
24	285.5	8.7	1464	12 US-10-060-036-159
25	285.5	8.7	1464	12 US-10-060-036-159
26	285.5	8.7	1464	12 US-10-060-036-159
27	285.5	8.7	1464	12 US-10-060-036-159
28	285.5	8.7	1464	12 US-10-060-036-159
29	285.5	8.7	1464	12 US-10-060-036-159
30	285.5	8.7	1464	12 US-10-060-036-159
31	285.5	8.7	1464	12 US-10-060-036-159
32	285.5	8.7	1464	12 US-10-060-036-159
33	285.5	8.7	1464	12 US-10-060-036-159
34	285.5	8.7	1464	12 US-10-060-036-159
35	285.5	8.7	1464	12 US-10-060-036-159
36	285.5	8.7	1464	12 US-10-060-036-159
37	285.5	8.7	1464	12 US-10-060-036-159
38	285.5	8.7	1464	12 US-10-060-036-159
39	285.5	8.7	1464	12 US-10-060-036-159
40	285.5	8.7	1464	12 US-10-060-036-159
41	285.5	8.7	1464	12 US-10-060-036-159
42	285.5	8.7	1464	12 US-10-060-036-159
43	285.5	8.7	1464	12 US-10-060-036-159
44	285.5	8.7	1464	12 US-10-060-036-159
45	285.5	8.7	1464	12 US-10-060-036-159

## ALIGNMENTS

RESULT 1  
US-09-770-517C-2  
Sequence 2, Application US/09770517C  
Patent No. US20020151035A1  
GENERAL INFORMATION:  
APPLICANT: Mitsuhashi, Kazuya  
APPLICANT: Yamamoto, Hiroaki  
APPLICANT: Matsuyama, Akiohbu  
APPLICANT: Tokuyama, Shinji  
TITLE OF INVENTION: D-AMINOACYLASE AND GENE ENCODING THE SAME  
FILE REFERENCE: 06501-072001  
CURRENT APPLICATION NUMBER: US/09/770,517C  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: JP 2000-019080  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: JP 2000-150578  
PRIOR FILING DATE: 2000-05-22  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 558  
TYPE: PRT  
ORGANISM: Hypomyces mycophilus  
US-09-770-517C-2  
Alignment Scores:  
Pred. No.: 2.4e-20  
Score: 612.00  
Length: 558  
Matches: 171

13

6.66e-07	Length:	558
310.00	Matches:	182

Percent Similarity: 35.07% Conservative: 27  
 Best Local Similarity: 30.54% Mismatches: 221  
 Query Match: 9,408 Indels: 167  
 DB: 9 Gaps: 34

US-10-009-782-1 (1-1758) x US-09-975-719-277 (1-558)

```

OY 60 CGACCTGCTGCTCCGGGGGCGGACCTTCAT-----CGACGGCAGCAACACCC 107
DB 34 ATGProsparglyArgGlyArgGlnProHisProHisArgAspArgLeuGlnAlaGly 53
OY 108 GGGGGCGGGCGGCGCTGGGGCGGCGGACCGCATCGGCCCATCGGCATCTGTC 167
DB 54 ThrGlySerSerArgProSerProAspProAlaGlnAspArgAlaArgGlyAlaGln 73
OY 168 GAGCGCGCGCGCGACAC-----CGGGGTGAGCTGTGGCGCTGCGGCGCC 218
DB 74 GLYArgArgArgSerHisGlnGlnAlaProGly---GlnAlaGlyGlyGlyTyrArgGln 92
OY 219 CGGCTTCATGAGCTCGGACACGACGACACTACCTGCTGCGGCGCGCA----- 272
DB 93 AlaArgAlaArgGlyLeuArgArgProAlaGlyAspLeuGlnValArgGlnGlyAla 112
OY 273 -----CATGACGCCAAGATCTGCGAGCGGCTGACGACGAGGTGTCACGGGCA 320
DB 113 GLYLeuGlyAlaAspProAlaGlnAspArgAlaGlyGlnAlaGlyArgGlyAla 132
OY 321 TTGCGGCAATCAGCTGCGCGCGCTGGCGGCGGACCGCGCGCGCGCGCTGACCTGCT 380
DB 133 AlaGlnGlyArgProArgGlnHisGlnAlaHisProValProAspHisProGlyProGly 152
OY 381 -----GAGCAAGCGGCTCTTACCGCTTTCAGCGGCTTCCGCGCACTACCT 425
DB 153 ThrGlnProAlaAspArgGlyArg-----ProAlaArgGlnAspArgGly----- 166
OY 426 GAGCGGCTTGGCGGCGGCGCGCGCGCTGACCGCGCGCTGATGGTGGCCATTCAC 485
DB 167 ---ProValAlaAlaGlnGlnGlnArgAspArgGlyAsnArgArgSerGly---PheGln 184
OY 486 GCTGCGCGC-----CGGCTCAGCGCGGAGCTTCCAGCGCGCGCGC 524
DB 185 ValAspArgTyrProGlyValGlnAspAlaArgGlyArgAlaArgGlnAlaAlaHis 204
OY 525 CACCGCAGGAAATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 584
DB 205 GLYAlaGlyAlaAlaSerAlaSerArgProGlyArgGly-----GlySerArgGly 222
OY 585 CATCGGCAATTCGACCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGAGAT 644
DB 223 -----ValGlnArgArgAlaProPheAlaArgProArgArgSerGlnProAla 239
OY 645 CATCGAGGTGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 704
DB 240 GlnArgLeuValArgProLeuPro----- 246
OY 705 CGAAGCGGAGCATCGTGGC-----CGGCTGGAGAGAACTTCGCG 746
DB 247 -----ArgProAspArgGlyGlyGlnAspArgValAlaGlnGlyAlaGlyArgVal 264
OY 747 CATCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 806
DB 265 LeuArgTyrArgGlyGly---AlaGlyAlaAspArgTyrValArgValHisGlyGlnThr 283
OY 807 CAATTTCGCGCGCTCGCGGAGAGCGTGGCGCT----- 839
DB 284 LeuGlyGlyProProAspArgArgAlaSerGlyLeuArgArgLeuArgGlyArgArgLeu 303
OY 840 ---GATCGAGCG-----CGCCATGGCGCGCGCGAGCGAGCTGCTCGCTGCG 878
DB 304 ProAspArgGlyAspProProGlnAlaLeuLeuGlyGlyAlaAlaGlyArg-----Gly 321
OY 879 CGCGATCCCAACGCGCGGCTCCACCATGCTCAGAGGAGCGCGCTGCTGCGCGCG 938
  
```

```

DB 322 GLYGlnGlyProSerGlyCysIleGlnHisSerProProGlyAlaArg-----GlyArg 339
OY 939 ACGCAACCATATCATCCTGGGCGCAAGCCCTTCCCGAAGTACGCGGCGGCGGCGGCGG 998
DB 340 ThrPro-----AspArgGlnSerArgAlaTyrGlyGlyLeu 351
OY 999 AGTCCGCGCGA---GCGCGGCAAAATCCAAAGTACGAGTGTGCGGCGGCGGCGGCGGCG 1055
DB 352 ProGlnHisArgGlyGlyAspArgPheGlnProArgPheGlyAlaAspProGlyAlaGly 371
OY 1056 CGGCGC-----CATCTACCTCATGATGAGCAAGCAAGCGGCGGCGGCGGCGGCGGCG 1103
DB 372 ArgArgProArgGlyAlaThrCysArgSerAspGlyArgGlyGlnCysAlaLeuProser 391
OY 1104 GCGGCT---GCGCGCGACAT-----GAT 1124
DB 392 GLYLeuHisGlnProAspArgArgArgSerGlyValArgAlaAlaGlySerArgAlaAsp 411
OY 1125 CGGCTCGGCGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1184
DB 412 ArgArgHisArgArgAspProAlaArgSerProAlaGlnAlaPro----- 426
OY 1185 GCGGCTGCTGGGCGCATATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1244
DB 427 -----GlnArgAlaArgAlaGlnProGlyThrGlnProGlyGlyAlaGly----- 441
OY 1245 GAAGATGACCGGCGCTGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1298
DB 442 -----GlnAlaAspCysArgArgLeuArgProGlyLeuThrProAlaGln 458
OY 1299 GCGAGCGCGG-----GTACTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1349
DB 459 AlaGlyHisProAlaLeuAspArgGlnProAlaGlyAlaThrAspProGly----- 475
OY 1350 CGATACCGGCGGCGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1394
DB 476 -----ArgGlnLeuArgAlaGlyCysGlnTyrLeuGlyGlyGlyArgArgArg 493
OY 1395 CCATTCCTGT-ACGTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1453
DB 494 ArgLeuArgLeuThrSer-----SerGlyAlaAspArgGlnSerProAlaSer 509
OY 1454 ATGCGCGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1513
DB 510 -----AlaGlyLeuPheHisGlyArgPro----- 517
OY 1514 GCGTAAAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1573
DB 518 -----ValGlyAlaArgAla----- 522
OY 1574 CTCCCTCCGCTCGCAATACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1621
DB 523 LeuProHisAlaValGlnGlyProPheCysArgSerTyrGlyGlyGly 538
  
```

RESULT 3  
 US-09-801-368-108  
 / Sequence 108, Application US/09801368  
 / Patent No. US20020128250A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Busby, Robert  
 / APPLICANT: Call, Brian  
 / APPLICANT: Hecht, Peter  
 / APPLICANT: Holtzman, Doug  
 / APPLICANT: Madden, Kevin  
 / APPLICANT: Maxon, Mary  
 / APPLICANT: Milne, Todd  
 / APPLICANT: No. US20020128250A1man, Thea  
 / APPLICANT: Royer, John  
 / APPLICANT: Salama, Sofie  
 / APPLICANT: Sherman, Amir  
 / APPLICANT: Silva, Jeff  
 / APPLICANT: Summers, Eric  
 / TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fung;

FILE REFERENCE: 109272.147  
 CURRENT APPLICATION NUMBER: US/09/801,368  
 CURRENT FILING DATE: 2001-03-07  
 PRIOR APPLICATION NUMBER: US 09/487,558  
 PRIOR FILING DATE: 2000-01-19  
 PRIOR APPLICATION NUMBER: US 60/160,587  
 PRIOR FILING DATE: 1999-10-20  
 NUMBER OF SEQ ID NOS: 440  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 108  
 LENGTH: 1367  
 TYPE: PRT  
 ORGANISM: Saccharomyces cerevisiae  
 US-09-801-368-108

Alignment Scores:  
 Pred. No.: 2.44e-06 Length: 1367  
 Score: 296.50 Matches: 139  
 Percent Similarity: 38.72% Conservative: 72  
 Best Local Similarity: 25.50% Mismatches: 266  
 Query Match: 8.99% Indels: 68  
 DB: Gaps: 13

US-10-009-782-1 (1-1758) x US-09-801-368-108 (1-1367)

QY 8 ACTGATGCGGAGAGAGATTTCATCGCATTCGATCCGATTCGACCTGC 67  
 DB 345 ThrSerSerThrThrluSerSerSerAlaProAlProHrProSerSerThrThr 364  
 QY 68 TGCCTCGCGCGCGACCTCATGACGCGACACCGCGCGCGCGCGCGCGCTGG 127  
 DB 365 GluSerSerAlaProAlProAlProAlProAlProAlProAlProAlProAlPro 382  
 QY 128 GCGTCGCGGCG 187  
 DB 383 ValThrSerSerThrThrluSerSerSerAlaProAlProHrProSerSerThrThr 398  
 QY 188 GCGTCG 238  
 DB 399 SerSerSerThrThrluSerSerSerAlaProAlProAlProAlProAlProAlPro 418  
 QY 239 -----CCACGACGACACTGCTGTCAGCGCGCGCGCGCGCGCGCGCGCGCG 292  
 DB 419 SerAlaProAlProHrProSerSerThrThrluSerSerSerAlaProAlProHrPro 426  
 QY 293 AGGCGCTCACCG 352  
 DB 437 SerThrThrluSerSerSerAlaProAlProAlProAlProAlProAlProAlPro 456  
 QY 353 CCAACCG 412  
 DB 457 ProAlProHrPro-----SerSerSerThrThrluSerSerSer 470  
 QY 413 TCGCCG-----ACTACGTGAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 460  
 DB 471 AlaProAlProHrProSerSerThrThrluSerSerSerAlaProAlProHrProSer 490  
 QY 461 CCGCGCTGATGCG 520  
 DB 491 SerThrThrluSerSerSerAlaProAlProAlProAlProAlProAlProAlPro 510  
 QY 521 CCGCGCGCGAGAGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 580  
 DB 511 ProAlProHrPro----- 514  
 QY 581 GCGCGATGCGCATTCGACCGCGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCG 640  
 DB 515 ---ProSerSerSerThrThrluSerSerSerAlaProAlProHrProSerSerSer 533  
 QY 641 AATATGATGAGGTGCG 700  
 DB 534 ThrThrlu-----SerSerSerAlaProAlProHrProSer 544

QY 701 GCGAGAGAGCGAGACATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 760  
 DB 545 SerThrThrluSerSerSerAlaProAlProHrProSerSerSerThrThrluSer 564  
 QY 761 TGGAGTCGCGGTGATCTCGACACCAAGATGATGGCGCGCGCGCGCGCGCGCG 820  
 DB 565 SerSerThrProAlProHrProSerSerThrThrluSerSerSerAlaProAlProHr 584  
 QY 821 CCGCGAGACCGTCCGCGTGAATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 880  
 DB 585 SerSerSerThrThrluSerSerSerAlaProAlProAlProHrProHrProSerSer 604  
 QY 881 CGTATCCCTACGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 940  
 DB 605 GluSerSerSerAlaProAlProAlProHrProSerSerSerThrThrlu-----Glu 620  
 QY 941 GCACCATCATCACT----- 976  
 DB 621 SerSerSerAlaProAlProAlProHrProSerSerThrThrluSerSerSerAlaPro 640  
 QY 977 TGAGCGGCG 1036  
 DB 641 ProSerSerSerThrThrluSerSerSerAlaProAlProAlProHrProHrProSerSer 660  
 QY 1037 TGCCGAGCTGACG 1096  
 DB 661 ThrGluSerSerSerAlaProAlProAlProHrProSerSerSerThrThrluSerSer 680  
 QY 1097 GCATCTGCGCGTGGCG-----CGACATGATCGCGCGCGCGCGCGCGCGCG 1150  
 DB 681 AlaProAlProHrProSerSerThrThrluSerSerSerAlaProAlProAlProHrPro 700  
 QY 1151 AGGCG 1210  
 DB 701 GluSerSerSerAlaProAlProAlProHrProSerSerSerThrThrluSerSerSer 720  
 QY 1211 ACCTGCGCGTGTCCGCTGAGACGCGCGGTATGAGATGACCGCGCGCGCGCGCG 1270  
 DB 720 aProAlProHrProHrProSerSerSerThrThrluSerSerSerAlaPro--ValProHr 739  
 QY 1271 GCTTGGCG 1330  
 DB 740 ProSerSerSerThrThrluSerSerSerAlaProAlProAlProHrProHrProHr 759  
 QY 1331 TGACCG 1390  
 DB 760 SerSerAlaPro--ValProHrProHrProSerSerSerThrThrluSerSerSerAla 778  
 QY 1391 GCATCCATTCGCTGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1450  
 DB 779 ValProHrProHrProSerSerSerThrThrluSerSerSerAlaProAlProHrPro 798  
 QY 1451 AGCATGCG 1510  
 DB 799 SerSerThrThrluSerSerSerAlaProAlProAlProHrProHrProSerSerSer 817  
 QY 1511 CCGCGGTAGCG 1570  
 DB 818 llerThrSerSerAlaProSerSerThrThrluSerSerSerThrThrluSerSerSer 837  
 QY 1571 CCCCTCCCT 1579  
 DB 838 ProAlPro 840

RESULT 4  
 US-10-025-380-1068  
 Sequence 1068, Application US/10025380  
 Publication No. US20020182191A1  
 GENERAL INFORMATION:  
 APPLICANT: Xu, Jianshun  
 APPLICANT: Lodes, Michael J.  
 APPLICANT: Secrist, Heather  
 APPLICANT: Benson, Darin R.



```

APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yuguang
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedrick, Thomas S.
APPLICANT: Carter, Darick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: OF COLON CANCER AND METHODS FOR THEIR USE
CURRENT APPLICATION NUMBER: US/10/025,380
NUMBER OF SEQ ID NOS: 1129
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1068
LENGTH: 5179
TYPE: PRT
ORGANISM: Homo sapiens
US-10-025-380-1068

```

```

Alignment Scores:
Pred. No.: 4.4e-06 Length: 5179
Score: 289.50 Matches: 146
Percent Similarity: 32.92% Conservative: 41
Best Local Similarity: 25.70% Mismatches: 164
Query Match: 8.78% Indels: 217
DB: 9 Gaps: 29
US-10-009-782-1 (1-1758) x US-10-025-380-1068 (1-5179)

```

```

QY 20 AAGAGAGATTTCAGTCCCAATCCGAT-----CCAGCCCTTCG 61
DB 1383 Lyslethrgalysnqscystpprometalsylscysilethrrhrproserpro 1402
QY 62 ACCTGCTGCTGGGGGGGACCCCTCATCGAGCGAGCAACACCCGGGGGGGGCGCG 121
DB 1403 ThrThrThrProserProProProThrThrThrThrThrPro-----Pro 1418
QY 122 ACCTGGGGCGGGGGGACCCCTCATCGAGCGAGCAACACCCGGGGGGGGCGCG 181
DB 1419 Thr-----ThrThrThrProserProProThrThrThrThrThrProPro 1434
QY 182 ACACCGGGGTGAGCTGTGCGGCTGTGCTGCGCGCGCGCTCATGACTCGACACCC 241
DB 1435 ThrThrThrPro-----SerProProlethrrhrthrthrthrPro 1448
QY 242 ACAGACACAACACTACGCTGCTCAGGCTGCGGACATGACGCCAAGATCTCGAGGGGTCA 301
DB 1449 LeuProThrThrThr-----Proser 1455
QY 302 CCACGGGTGATACGGGCAATGCGGCATACGCTGCGCGCGCGTGGCGACGCCACCGC 361
DB 1456 ProProleSer-----ThrThrThrThr 1463
QY 363 CGCCCCCTGGACCTGTGAGCAAGAGGGGCTTACGCTTTCAGCGCTTGGCGACT 421
DB 1464 ProProProThrThr-----ThrProserProPro 1473
QY 422 ACCTGAGCGCTGGGGGGGACCGCGCGCGCTGACACGCCCTGTATGTGGGCGCAT 481
DB 1474 ThrThrThrProserProProThrThrThrProserProProThrThr 1489
QY 482 CAAGCTGGGGCGGGGATGCGGACTTCAGAGGGCGCCCGACCGAGAGAATCG 541
DB 1490 -----ThrThrThrThrProProProThrThrThrPro 1500
QY 542 CGGCATCGGGACCTGGCGCGAGAGCATGGCAGGGCGCCATTCGCAATTGACCG 601
DB 1500 -----ThrThrThrThrProProProThrThrThrPro 1500

```

```

DB 1501 SerProProleThrThrThrThrPro-----Pro 1511
QY 602 GGGCTTCTACCCGCCCGCGCGCGCCACCGAGAGATCATGAGGTGCGCGC 661
DB 1512 AlaserThrThrThrThrProProThrThrThrPro----- 1523
QY 662 CGCTAGCGCGCATGCGGCATCTACGACCCACATCGGGAGAGAGCGACATCG 721
DB 1524 -----SerProProThrThrThr-----ThrThr 1532
QY 722 TGGCGCGCTGAGGAACCTTCGCGATCGCGCGCGAGCTGCGGTGATCT 781
DB 1533 ProProProThrThrThrProserProProThr----- 1543
QY 782 CGACACCAAGGTATGGGCCAATTCGCGCGCGCGCGAGAGCTGCGCGTGA 841
DB 1544 -----ThrThrProleThr----- 1548
QY 842 TCGAGCGCGCATGCGCGCGCGCGCGCTGCTGAGCGGTATCTCT--ACGTGGCG 898
DB 1549 -----ProPro-----ThrThrThrThrThrThrThrThrThrThr 1562
QY 899 GCTCACCATGCTCAAGACGACCGCGCTGCTGCGCGCGAGCAGCATCATCTGT 958
DB 1563 SerProProProThrThrThrThrThr-----ProProThr 1575
QY 959 GCAAGCCCTTCCCGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1018
DB 1576 ThrThrThrProserPro-----ProThrThrThr 1584
QY 1019 AATCCAGT-----ACGACGTGTGCGCGAGCTGACCGCGCGCGCGCA 1063
DB 1585 ThrProserProProThrThrThrThrThrThrThrThrThrThrThrThrThr 1604
QY 1064 TCTACTCATGATGAGGAAC-----CGACGTGAGCGCATCTGCGTGGCCCA 1117
DB 1605 ProthrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1624
QY 1118 CATGATGCGCTCGAGCGCGCTGCGCGCGAGAGCGCGCGCATCGCGCGTGGGCA 1177
DB 1625 ProlethrProProThrSer-----ThrThr----- 1633
QY 1178 CMTCCCGCGGTGCTGGGCACTATGCGCGGAGCTGCGCTGCTCCGCTGAGACGG 1237
DB 1634 -----ThrleuProProThrThr----- 1640
QY 1238 CGGTATGAGATGACCGCGCTGACCGCGCGCGCTTCCGCGCGCGCGCGCGAC 1297
DB 1641 -----ProserProProProThrThrThrThrThrThrThrThrThrThr 1651
QY 1298 TGCAGCGCGGTACTGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGATA--- 1354
DB 1652 --ProProThrThrThrProserProProThrThrThrThrThrThrThrThrThr 1670
QY 1355 -----CGCGACCTTCGAGACCCCTACCGAGCGCGCG----- 1387
DB 1671 ThrThrThrThrProProProThrThrThrProserProleThrThrThrProser 1690
QY 1388 -----CGGCAATTCATTCGGT 1405
DB 1691 ProProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1709
QY 1406 ACCTACAGCGCGCGCGCTGCGAGAGACGCGCTGACCGCGCGAGCATCGCGCGCG 1465
DB 1710 ThrThrThrThrThrThrThr-----SerThrThrThrThrThrThrThrThrThr 1727
QY 1466 TGCCTGACGACGCGCGCGCTGAGCGCGCGCGCGCGCGCTTTCATTCGCGCGTGAACGGGG 1535
DB 1728 MetThr-----ThrProserProThrThrThrThrThrThrThrThrThrThr 1744
QY 1536 CGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1585
DB 1745 ThrThrleuProProThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1763

```



Db 1710 ThrThrThrThrThrProser-----SerThrThrThrProserProProThrThr 1727  
 Oy 1466 TCGTGCAGCAGCGGCGGCTGAGCCCGGCGGCGGCTTACATCCGCGGAGAGGG 1525  
 Db 1728 MetThr-----ThrProserProThrThrThrProserProThrThrThrMet 1744  
 Oy 1526 CGGCGTGGCGCCCGCTCCCAACCTGAGCGCAACCGGTACATGGCCCTCCCTCCGCTC 1585  
 Db 1745 ThrThrLeuProThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1763  
 Oy 1586 GCATACGCGCCCGCCACCATATCG 1609  
 Db 1764 SerThrThrProThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1771

## RESULT 6

US-09-833-263-1068  
 : Sequence 1068, Application US/09833263  
 : Patent NO. US20020110547A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Wang, Aijun  
 : APPLICANT: Clapper, Jonathan D.  
 : APPLICANT: Stolk, John A.  
 : APPLICANT: Meagher, Madeline J.  
 : TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
 : FILE REFERENCE: 210121.471C12  
 : CURRENT APPLICATION NUMBER: US/09/833,263  
 : NUMBER OF SEQ ID NOS: 1093  
 : SOFTWARE: FastSeq for Windows Version 3.0  
 : SEQ ID NO 1068  
 : LENGTH: 5179  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 : US-09-833-263-1068

## Alignment Scores:

Pred. No.: 4.4e-06 Length: 5179  
 Score: 289.50 Matches: 146  
 Percent Similarity: 32.92% Conservative: 41  
 Best Local Similarity: 25.70% Mismatches: 164  
 Query Match: 8.78% Indels: 217  
 Db: 10 Gaps: 29

US-10-009-782-1 (1-1758) x US-09-833-263-1068 (1-5179)

Oy 20 AAGAGAGATTCATGTCATCCATCCGAT-----CCACGCCCTTCG 61  
 Db 1383 LysIleArgValAsnGlySerProMetAspLysCysAlaThrThrProserProPro 1402  
 Oy 62 ACCGTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 121  
 Db 1403 ThrThrThrThrProserProProThrThrThrThrThrThrThrThrThrThrThr 1418  
 Oy 122 ACCGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 181  
 Db 1419 Thr-----ThrThrProserProProThrThrThrThrThrThrThrThrThrThrThr 1434  
 Oy 182 ACACCGGGGTGACGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241  
 Db 1435 ThrThrThrPro-----SerProThrThrThrThrThrThrThrThrThrThrThr 1448  
 Oy 242 ACAGCAGACACTACCTGCTCAGGCGTGCAGACAGCCCAAGATCGCAGGCGCTCA 301  
 Db 1449 LeuProThrThrThr-----Proser 1455  
 Oy 302 CCACGGTGTGACGGGGAATGGCGCATCAGCTGCGCCCGCTGGCGCAGCCCAACCGC 361  
 Db 1456 ProProThrThr-----ThrThrThrThr 1463  
 Oy 362 CCGCGCCCGCTGACCTGCTGAGCAGAGCGGCGCTTACGTTTGAGCGCGTTCCGCGACT 421  
 ||||| |||||

Db 1464 ProProThrThr-----ThrProserProPro 1473  
 Oy 422 ACCGTGAGCGGTGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 481  
 Db 1474 ThrThrThrThrProserProProThrThrThrThrThrThrThrThrThrThrThr 1489  
 Oy 482 CAAGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 541  
 Db 1490 -----ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1500  
 Oy 542 CGGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 601  
 Db 1501 SerProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1511  
 Oy 602 GCGCCTTACCG 661  
 Db 1512 AlaSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1523  
 Oy 662 CGGTGAGCGGCGATGGCGGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 721  
 Db 1524 -----SerProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1532  
 Oy 722 TGGCGGCGGTGAGGAACCTTCGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 781  
 Db 1533 ProProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1543  
 Oy 782 CGCAGCACAAGTATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 841  
 Db 1544 -----ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1548  
 Oy 842 TCGAGCGCGCGATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 898  
 Db 1549 -----ProPro-----ThrSerThrThrThrThrThrThrThrThrThrThrThr 1562  
 Oy 899 GCTGCACATGCTCAGACGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 958  
 Db 1563 SerProProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1575  
 Oy 959 GCAGCCCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1018  
 Db 1576 ThrThrThrThrPro-----ProThrThrThr 1584  
 Oy 1019 AATCCAGT-----ACGAGTGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1063  
 Db 1585 ThrProserProProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1604  
 Oy 1064 TCTACTCATGATGAGACAGAC-----CCGAGTGCAGCGCATCTGCGTGGCGGCGGCGGCGG 1117  
 Db 1605 ProThr 1624  
 Oy 1118 CCATATCGGCTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1177  
 Db 1625 ProThr 1633  
 Oy 1178 COTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1237  
 Db 1634 -----ThrLeuProProThrThrThr 1640  
 Oy 1238 CGGTATGGAATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1297  
 Db 1641 -----ProserProProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1651  
 Oy 1298 TGCAGCGCGGCGTACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1354  
 Db 1652 -----ProProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1670  
 Oy 1355 -----CCGCGACTTGGACACCTACGAGCGGCGG----- 1387  
 Db 1671 ThrThrThrThrProProProProThrThrThrThrThrThrThrThrThrThrThrThrThr 1690  
 Oy 1388 -----CCGAGTCACCATTCGCTG 1405  
 Db 1691 ProProThr 1709

QY 1406 ACCTGACGCGCGCGCTGTGGCAAGAGAGCGGCTTACCGCGCCAGATGCCGCGCG 1465  
 Db 1710 ThThrThrThrThrProser-----SerThThrThrProserProProThThr 1727  
 QY 1466 TGCCTGACGACGCGCGCGCTGTAGCCGCGCGCGCTTACATCGCGCGTAAAGGGG 1525  
 Db 1728 MetThr-----ThThrProserProThThrThrProserProThThrMet 1744  
 QY 1526 CGCGGTGCGCGCGCGCTTCCACACCTGTAGCGCAACCGCTTACGCGCGCTCCGCGCTC 1585  
 Db 1745 ThThrLeuProThThrThrThrSerSerProLeuThThrThrProLeuProPro---- 1763  
 QY 1586 GCATATAGCGCGCGCGCGATATCG 1609  
 Db 1764 SerLeuThrProThThrPheSer 1771

## RESULT 7

US-10-124-557-14

Sequence 14, Application US/10124557  
 Patent No. US20020137894A1  
 GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Gesner, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Geert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-5851

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 941 amino acids

TYPE: amino acid

MOLECULE TYPE: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-10-124-557-14

Alignment Scores:

Pred. No.: 7,81e-06 Length: 941

Score: 285.50 Matches: 137

Percent Similarity: 33.508  
 Best Local Similarity: 23.308  
 Query Match: 8.658  
 DB: 12  
 Gaps: 24

US-10-009-782-1 (1-1758) x US-10-124-557-14 (1-941)

QY 32 CCATGCCCAATCCGATTCCAGCCCTTGACCTGTGCTCGCGGCGCACCTCATCG 91  
 Db 167 ProThrProThrThrThrLeysSerAlaProThrThrProLysGluProAlaProThr 186  
 QY 92 ACGCAGCA-----ACACCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142  
 Db 187 ThThrsSerAlaProThrThrProLysGluProAlaProThr-----Thr 201  
 QY 143 GCATCGCGCCATCGCGCATCTGTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 202  
 Db 202 ThThrsGluProAlaPro-----ThThrProLysGluProAlaProThr----- 216  
 QY 203 GCGTGTGTGTGCG 262  
 Db 217 -----ThThrThrsGluProAlaProThrThrThrsGluProAlaProThrThr 228  
 QY 263 GCGCTCGCATATGACCG 322  
 Db 229 AlaProThrThr-----ProLysGluProAlaPro----- 238  
 QY 323 GCGGCATACGCTGCG 382  
 Db 239 -----ThThrThrsGluProAlaProThrThrThrsGluProAlaProThrThr 442  
 QY 383 ACGAAGCGGCTCTTACGTTGAGCGGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG 442  
 Db 251 GluProAlaProThrThrThrProLysGluProThrProThrThrProLysGluProAla 270  
 QY 443 CGCGCGCGG-----CGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493  
 Db 271 ThThrThrsGluProAlaProThrThrPro----- 280  
 QY 494 CGCGGTGATGCG 553  
 Db 281 -----LysGluProAlaProThrAlaProLysLysProAlaPro 293  
 QY 554 ACCTGCGCGAAGGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 613  
 Db 294 ThThrThrsGluProAlaProThrThrThrProLys-----GluProAlaProThr 311  
 QY 614 CG 670  
 Db 312 ThThrsGluProSerProThrThrThrProLysGluProAlaProThrThrThrsSerAla 331  
 QY 671 CGCATGGGGGATCTACCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730  
 Db 332 ProThrThrThrThrsGluProAlaProThrThrThrThrsSerAlaProThrThrProLys 351  
 QY 731 TGGAGGAACCTTCCGATCG 790  
 Db 352 GluProSerProThr-----ThThrThrsGluProAlaProThrThr 365  
 QY 791 AGGTCATGGCG 850  
 Db 366 ProLysGluProAlaPro-----ThThrPro 374  
 QY 851 CCATGGCG 910  
 Db 375 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 394  
 QY 911 TCAAGCAGACGCGGTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970  
 Db 395 ProThrThrThrsLysLysProAlaProThrAlaProLysGluProAlaProThrThrPro 414  
 QY 971 CCGAATGACG 1030

```

Db 415 LysGluThrAlaProThrProLysLysLeuThrProThrThrProGluLysLeuAla 434
QY 1031 ACCTGGTCCCGGAGCTGACGCGCGGCGGCACTACTCATATGATGAGACACCCGAG 1090
Db 435 ProThrThrPro-----GluLysProAlaProThrThrProGluLysLeuAlaProThr 452
QY 1091 TGCAGCGCATCTGGGCTTGGCCGCGGCACTGATGCG--GCTCCGAGCGGCTGCGGACG 1147
Db 453 ThrProGluLupProThrProThrThrProGluLupProAlaProThrThr----- 469
QY 1148 ACAGAGGCGCGCATCCGCGCTGTGGGACACTTCCCGGGGTGGGCGACTATGCGC-1207
Db 469 ----- 469
QY 1208 GCGACCTGGGCTGTTCCTGCGGAGACGCGGATGAGATGACCGGCTGACCCCG 1267
Db 470 -----Pro 470
QY 1268 CCGCTGCGGCTGCGGCGCGGCGGCGGAGCTGCGAGCGGCTGCTGCGGAGCTGTG 1327
Db 471 LysAlaAlaAla---ProAnThrProLysGluProAlaProThrThrProLysGluPro 489
QY 1328 TGTTCACCCCGGCGGAGGTTGGCCGATACG---CCACCTTGAAACACCTACCGAGCG 1384
Db 490 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 509
QY 1385 CCGCGCGCATCATCTGCTGATGCTACAGCGCGCGCGGCTGCTGCGAGACGCGCTCA 1444
Db 510 ProLysGluThrAlaPro---ThrThrLeuLysGluProAlaProThrThrProLysLys 528
QY 1445 CCGCGCGCATCGCGCGCGGCTGCTGCGAGCGAGCGCGCGGCGCGGCGGCGCGCGCT 1504
Db 529 ProAla-----ProLysLysLeuAlaProThrThrThrLysGluProThrSerThr--- 545
QY 1505 TACATCCGCGCTGAACGGGCGGCGGCGGCGGCGGCGGCTGCCAACCTGAGCAACCGCT 1564
Db 546 -----TherSerLysProAlaProThrThrThrProLysGluThrAlaPro 560
QY 1565 ACATGCGCCCTCCCTCCGCTGCGAATAGCGCGCGCCACCATATCGTGGGCAAGAAATGA 1624
Db 561 ThrThrProLysGluProAlaProThrThrPro----- 571
QY 1625 TGGGCGCGGCTGCGGCGGCGGAGCGCGGAAATGACCTGCAAGACCTGCGC 1684
Db 572 -----LysGluProAlaProThrThrProLysGluThrAlaPro---ThrThrLeuLys 588
QY 1685 AGCGGAGG---GCAATCGCGGCTGCGACCTGTCCAGG-----CCGAGCTGGGCGC 1732
Db 589 GluProAlaProThrThrProLysLysProAlaProLysGluLeuAlaProThrThrThr 608
QY 1733 AGATGCGCCCTGAGTACGAGGAGC 1756
Db 609 LysGluProThrSerThrThrSer 616

RESULT 8
US-10-124-557-84
; Sequence 84, Application US/10124557
; Patent NO. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Csegr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84

Alignment Scores:
Pred. No.: 7,75e-06 Length: 1022
Score: 285.50 Matches: 137
Percent Similarity: 33.50% Conservative: 60
Best Local Similarity: 23.30% Mismatches: 240
Query Match: 8.65% Indels: 151
Gaps: 24

US-10-009-782-1 (1-1758) x US-10-124-557-84 (1-1022)
QY 32 CCATGTCCTCAATCCGATTCACGCTTCGACCTGCTGCGGCGGCGGCGGCGGCGGCGG 91
Db 248 ProThrProThrThrThrLysSerAlaProThrThrProLysGluProAlaProThrThr 267
QY 92 ACGGCAACA-----ACACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 142
Db 268 ThrLysSerAlaProThrThrProLysGluProAlaProThr-----Thr 282
QY 143 GCATGCGCGCATCGGCGGATCTGTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 202
Db 283 ThrLysGluProAlaPro-----ThrThrProLysGluProAlaProThr----- 297
QY 203 GCGTGTGTGTCGCGCGGCTTCATGACGATCGACACCGACGACGACGATGCTGCTCA 262
Db 298 -----ThrThrLysGluProAlaProThrThrThrLysSer 309
QY 263 GCGGTGCGCATGACGCGCCCAAGATCTGACAGGCGGTCACCAAGGTGTGTCAGGCGCAATT 322
Db 310 AlaProThrThr-----ProLysGluProAlaPro----- 319
QY 323 GCGGATCATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 382
Db 320 -----ThrThrProLysLysProAlaProThrThrProLys 331
QY 383 ACGAAGGCGGCTTACCGCTTCGAGCGCTTCGCGGATCTGAGACGCGTTCGGGCGCA 442
Db 332 GluProAlaProThrThrProLysGluProThrThrThrProLysGluProAlaPro 351
QY 443 CCGCGGCGG-----CGTCAACGCGCGCTGTATGATGGCGCATTCACGCTGCGG 493

```

```

Db 352 ThrThrLysGluProAlaProThrThrPro----- 361
OY 494 CCGCGGCTATGCGGAGCTTGCAGCGCGCGGACGAGAGAAATCCGGCCATCGCGG 553
Db 362 -----LysGluProAlaProThrAlaProLysGluProAlaPro 374
OY 554 ACCGCGCGAGAGAACCTATGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 613
Db 375 ThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 392
OY 614 CGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 670
Db 393 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrThrLysSerAla 412
OY 671 CGCATGGCGGCGCATCTACGCCACCCACATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 730
Db 413 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 432
OY 731 TGGAGAAACCTCCCGCATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 790
Db 433 GluProSerProThr-----ThrThrLysGluProAlaProThrThr 446
OY 791 AGGTATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 850
Db 447 ProLysGluProAlaPro-----ThrThrPro 455
OY 851 CCATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 910
Db 456 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 475
OY 911 TCAAGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 970
Db 476 ProThrThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrThrPro 495
OY 971 CCGAATAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1030
Db 496 LysGluThrAlaProThrThrProLysLysLeuThrProThrProGluLysLeuAla 515
OY 1031 ACGTGTGCGGAGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1090
Db 516 ProThrThrPro-----GluLysProAlaProThrThrProGluLysLeuAlaProThr 533
OY 1091 TGCAGCGCATCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1147
Db 534 ThrProGluLysProThrProThrProGluLysLeuAlaProThr----- 550
OY 1148 ACGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1207
Db 550 ----- 550
OY 1208 GCGAGCTGGGCGTGTCCCGCTGAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1267
Db 551 -----Pro 551
OY 1268 CCGGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1327
Db 552 LysAlaAlaAla---ProsnThrProLysGluProAlaProThrThrProLysGluPro 570
OY 1328 TGTGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1384
Db 571 AlaProThrThrProLysGluProAlaProThrThrProLysGluAlaProThrThr 590
OY 1385 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1444
Db 591 ProLysGluThrAlaPro---ThrThrLeuLysGluProAlaProThrThrProLysLys 609
OY 1445 CCGGCGGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1504
Db 610 ProAla-----ProLysGluLeuAlaProThrThrThrLysGluProThrSerThr--- 626
OY 1505 TACAATCCGGCGTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1564

```

```

Db 627 -----ThrSerAspLysProAlaProThrThrProLysGluThrAlaPro 641
OY 1565 ACATGCGCCCTCCCTCCGCTCGCATACGCGCCACCGCATATGCTGGCGAAGAGTGA 1624
Db 642 ThrThrProLysGluProAlaProThrThrPro----- 652
OY 1625 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1684
Db 653 -----LysGluProAlaProThrThrProLysGluThrAlaPro---ThrThrLeuLys 669
OY 1685 AGGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1732
Db 670 GluProAlaProThrThrProLysLysProAlaProLysGluLeuAlaProThrThrThr 689
OY 1733 AGATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1756
Db 690 LysGluProThrSerThrThrSer 697

```

RESULT 9  
 US-10-124-557-74  
 Sequence 74, Application US/10124557  
 Patent No. US20020137894A1  
 GENERAL INFORMATION:  
 APPLICANT: Turner, Katherine  
 Clark, Stephen C.  
 Jacobs, Kenneth  
 Hewick, Rodney M.  
 Gesner, Thomas G.  
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 NUMBER OF SEQUENCES: 143  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genetics Institute, Inc.  
 STREET: 87 Cambridgepark Drive  
 CITY: Cambridge  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02140

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/124,557  
 FILING DATE: 16-Apr-2002  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/643,502  
 FILING DATE: 18-JAN-1991  
 APPLICATION NUMBER: US 07/546,114  
 FILING DATE: 29-JUN-1990  
 APPLICATION NUMBER: US 07/457,196  
 FILING DATE: 29-DEC-1989  
 APPLICATION NUMBER: US 07/390,901  
 FILING DATE: 08-AUG-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Cseert, Luann  
 REGISTRATION NUMBER: 31,822  
 REFERENCE/DOCKET NUMBER: GI 5190  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)876-1170  
 TELEFAX: (617)876-5851  
 INFORMATION FOR SEQ ID NO: 74:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1038 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 74:

Alignment Scores: 7.73e-06 Length: 1038  
 Pred. No.:



OY	32	CGATGTCGACCAATCCGATTCCGACACCCTTCGACTGTGTGCGGGGGGACACCTCATCG	91
Db	232	ProthrProthrThrLlysSerAlaProthrThrProLysGluProAlaProthrThr	251
OY	92	ACGGCAGCA-----ACACCCGGGGGGCGCGCGACCTGGCGCTGGCGGGCAGC	147
Db	252	ThrLysSerAlaProthrThrProLysGluProAlaProthr-----Thr	266
OY	143	GCATCCCGCGCATCGCGCATCTGTGGACGCCCGCGCACACCGGGGTGACGTTCGG	207
Db	267	ThrLysGluProAlaPro-----ThrThrProLysGluProAlaProThr-----	281
OY	203	GCCGTGTGTGCGCGCCGCTTCATCGACTCGACACCCAGACAGCAACACTGCTCA	265
Db	282	-----ThrThrLysGluProAlaProthrThrLysSer	293
OY	263	GGCGTGGCACAATGACGCCCAAGATCTGCAGAGGGCTCACACGAGTGTACGGGCAATT	322
Db	294	AlaProthrThr-----ProLysGluProAlaPro-----	303
OY	323	GGGCATCAGCTGGCGCGCGTGGGACAGCCAAACCGCCCGCCCTTGACTGCTGG	382
Db	304	-----ThrThrProLysLysProAlaProthrThrProLys	315
OY	383	ACGAAGCGGCTTACGCTTTCGAGCGCTTCGCGCATCTGACGAGCGTGGGGGCA	442
Db	316	GluProAlaProthrThrProLysGluProthrProthrThrProLysGluProAlaPro	335
OY	443	CGCGGGCGG-----CGGCAAGCGCGCGCTTAAGTGGCCATTCAACCTGGCG	493
Db	336	ThrThrLysGluProAlaProthrThrPro-----	345
OY	494	CGCGGTCATGCGGACCTTCGACGCGCGCGCCACGACGAGAAATCGGGCATGGGG	553
Db	346	-----LysGluProAlaProthrThrAlaProLysLysProAlaPro	358
OY	554	ACCTGGCGGAGAAACCATGGCCACGCGCGCCCATCGGCAATTGACGCGGCTTAC	613
Db	359	ThrThrProLysGluProAlaProthrThrProLys-----GluProAlaProthrThr	376
OY	614	CGCCGCGCGCGGCGCA--CGACGAAGAAGATCATCGAGGTGCGCGCGCTGAGCG	670
Db	377	ThrLysGluProSerProthrThrProLysGluProAlaProthrThrThrLysSerAla	396
OY	671	CGCATGGCGCATCTAGCGCACCCACATGCGCGACGAGAGCGACACATCTGGCGCGC	730
Db	397	ProthrThrThrLysGluProAlaProthrThrThrLysSerAlaProthrThrProLys	416
OY	731	TGGAGAAACTTCCGATCGCGCGCGAGCTGAGCTGCGCGGTGTGATCTGCACACCA	790
Db	417	GluProSerProthr-----ThrThrLysGluProAlaProthrThr	430
OY	791	AGGTCATGGGCGACCAATTTGGCGCGCTCGCGCGCAGAGCGTCCGCTGATCGAGCGC	850
Db	431	ProLysGluProAlaPro-----ThrThrPro	439
OY	851	CCATGGCGCGCAGAGCTCTCGCTGAGCAGCGATCCCTACGAGCGCGCTCCACATGC	910
Db	440	LysLysProAlaProthrThrProLysGluProAlaProthrThrThrProLysGluProAla	459
OY	911	TCAAGCAGAGACCGGTCTCTGCGCGGAGCAGCACATCATCACTGTGTCAAGCCTTCC	970
Db	460	ProThrThrThrLysLysProAlaProthrAlaProLysGluProAlaProthrThrPro	479
OY	971	CGGAATGACGCGGGCGGACCTGGATGAAGTGGCGGCGGACGCGGCAATTCAAATGAC	1030

Db 480 LysGluThrAlaProThrThrProLysLysLeuThrProThrThrProGluLysLeuAla 49  
 1031 ACCTGGTGGCCGAGCTGACGACCGCGGGCCATCCTACTCATATGATGACGAACCCGACG 10  
 Db 500 ProThrThrPro-----GluysPProAlaProThrThrProGluLysLeuAlaProThr 51  
 1091 TGCAGCGCATCTCGCGGTTCCGGCCGACCATGATCG--GCTCCGACGCGCTGCGCGACG 11  
 Db 518 ThrProGluLysProThrProThrProThrProGluLysProAlaProThrThr----- 53  
 1148 ACGAGCGCCCGCATCCGCGCCCTGTGGGGACCTCCCGCGGGTCTGGGGCATTATGCGC 12  
 Db 534 ----- 53  
 1208 GCGACCTGGCGCTGTTCCTCCGCTGAGACGCGCGTATGGAATGACGCGGCTGACCGCGG 12  
 Db 535 -----Pro 53  
 1268 CCGCGTTCGCGCTGGCGCGCGCGCGCGACCTGACGCGCGGACTTCGCGCATCTGTGG 13  
 Db 536 LysAlaAlaAlaAla---ProAsnThrProLysGluProAlaProThrThrProLysGluPro 55  
 1328 TGTTCGACCGCGCGACGAGGGCGCGATACG---CCACTTGAACACCTTACCGAGCGG 13  
 Db 555 AlaProThrThrProLysGluProAlaProThrThrProLysGluLysAlaProThrThr 57  
 1385 CCGCGCGCATTCCTATTCGTACTGATCAACAGCGCGCGCGCTGTCGCAAGAGCGCGCTCA 14  
 Db 575 ProLysGluThrAlaPro---ThrThrLeuLysGluProAlaProThrThrProLysLys 59  
 1445 CCGCGCGACAGTCGCGCGCGCGCTGCTGCACAGCAGCGCGCGCTGAGCCCGCGCGCGCT 15  
 Db 594 ProAla---ProLysGluLeuAlaProThrThrThrLysGluProThrSerThr--- 61  
 1505 TACAATCCGCGGTGATACGCGCGCGCGGTGCGCGCCCGCCCGCAACCTGAGCCAAACGCT 15  
 Db 611 -----ThrSerAspLysProAlaProThrThrProLysGluThrAlaPro 63  
 1555 ACATGCGCCCTCCCTCCGCTGCGCATACGCGCCGACCCGATATCGTGGCAAGAGATGA 16  
 Db 626 ThrThrProLysGluProAlaProThrThrPro----- 63  
 1625 TGGGCGCGCGCTGGCGCGCGAGGCGCAAGCGCCGGAAATGACCTCTCAAGACCTGTGCG 16  
 Db 637 -----LysGluProAlaProThrThrProLysGluThrAlaPro---ThrThrLeuLys 65  
 1685 AGGCCAGCG---GCATCGCGGCTCGACCCCTGTCAGG-----CCGACTGTGGC 17  
 Db 654 GluProAlaProThrThrProLysLysProAlaProLysGluLeuAlaProThrThr 67  
 1733 AGATCGCCCTGAGCTACGAGAAAGC 1756  
 Db 674 LysGlyProThrSerThrThrSer 681

RESULT 10  
 US-10-124-557-58  
 Sequence 58, Application US/10124557  
 Patent No. US20020137894A1  
 GENERAL INFORMATION:  
 APPLICANT: Turner, Katherine  
 Clark, Stephen C.  
 Jacobs, Kenneth  
 Hewick, Rodney M.  
 Gessner, Thomas G.  
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 NUMBER OF SEQUENCES: 143  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genetics Institute, Inc.  
 STREET: 87 Cambridgepark Drive  
 CITY: Cambridge  
 STATE: Massachusetts  
 COUNTRY: U.S.A.

```

QY 32 CCAATGCGCAACCGATTCCAGACCCCTTGACCTGCTCGAGGGGAGCAACCCATCG 91
Db 275 ProthPrProthThrThrIleYSerAlProThrThrProlySgluProAlaProThrThr 295
QY 92 ACGGCA-----ACACCGGGGGGGGGCGGACCTGGGCGTGGCGGAC 147
Db 295 ThrYSerAlProThrThrThrProlySgluProAlaProThr-----Thr 309
QY 143 GCATGCGCCGATCGGCGGATGTCGAGAGCGCGCGCGGACACCCGGGTGACGTGCG 202
Db 310 ThrYSgluProAlaPro-----ThrThrProlySgluProAlaProThr----- 324
QY 203 GCGTGTGTGCGCGCGCGGCTTCATCGACTGCGACCGACCGACAACTACCTGCTCA 262
Db 325 -----ThrThrYSerAlProAlaProThrThrThrYSer 336
QY 263 GCGCTCGGACATACGCCCAAGATCTGCGAGGGCGGACACCAAGGTGTGACGGGAATT 322
Db 337 AlaProThrThr-----ProlySgluProAlaPro----- 346
QY 323 GCGGCAACCGCTGGCGCGCTGCGGCGACGCCAACCGCGCGCGCGCGGACCTGTGG 382
Db 347 -----ThrThrProlySgluProAlaProThrThrProlyS 358
QY 383 ACGAGCGGCGCTTACCGCTTTCGACCGCTTTCGCGGACTACCTGAGACGGTGGCGCA 442
Db 359 GluProAlaProThrThrProlySgluProAlaProThrThrThrProlySgluProAlaPro 378

```

OY 444 CGCGCGCGG-----CGTCAACGCCGCTGTATGTGGCCATTCAACGCTGC CG 497  
 Db 379 ThrThrLysGluProAlaProThrThrPro-----ThrPro----- 386  
 OY 494 CGCGGTCATGCGGACACTTGCAGCGCGCCGCCACCGACGAGAAATCGCGCCATGCGG 555  
 Db 389 -----LysGluProAlaProThrAlaProLysProAlaPro 401  
 OY 554 ACCTGGCCGAGAGACCTATGTGACAGCGCGCCCATCGGCACTTTCAGCCGCGCTTAC 613  
 Db 402 ThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 419  
 OY 614 CGCGCGCGCGCGCGCA---CCACGAGAGAATCATGAGTGTGCGCGCGCTGACG 670  
 Db 420 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrThrLysSerAla 439  
 OY 671 CGCATGGCGGCATATACGCCACCCACATGCGCGCAGAGCGACATCGTGGCGCG 730  
 Db 440 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 459  
 OY 731 TGGAGGAACCTTCCGCATCGCGCGGAGCGTGCAGCGCGGCGGATCTCGACACCA 790  
 Db 460 GluProSerProThr-----ThrThrLysGluProAlaProThrThrThr 473  
 OY 791 AGGTCATGGCGCCCAATTTCGGCCCTCGCGCGAGACGCTGCGCGTGAACGAGCG 850  
 Db 474 ProLysGluProAlaPro-----ThrThrPro 482  
 OY 851 CCATGGCGCGCGAGACGTCCTGCTGCGTGCAGCGGATCCCTACGTGCGCGCTCCACCAT 910  
 Db 483 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 502  
 OY 911 TCAAGAGACCGCGCTGCTGCTGCGCGAGCACCATCATGACCTGCGTGAAGGCCCTTC 970  
 Db 503 ProThrThrThrLysLysProAlaProThrThrAlaProLysGluProAlaProThrPro 522  
 OY 971 CGCACTGAGCGGGGCGCGACCTGGATGAATGCGCGCGCGAGCGCGGCAATCATCAATG 1030  
 Db 523 LysGluThrAlaProThrThrProLysLysLeuThrProThrThrProGluLysLeuAla 542  
 OY 1031 ACGTGTGCCGAGCTGCAGCGCGCGCGCGCATCTATCATATGAGACGACCGCGAG 1090  
 Db 543 ProThrThrPro-----GluLysProAlaProThrThrProGluLysLeuAlaProThr 560  
 OY 1091 TGCAGCGCATCTGGCGCTGGCGCGCACATGATCG---GCTCGAGCGCGCTCGCGAG 1147  
 Db 561 ThrProGluThrProThrProThrThrProGluLysProAlaProThrThr----- 577  
 OY 1148 ACGAGCGCGCGCATCGCGCGCTGGGGGACACTTCCGCGGCTGGGGGCACTATGCG 1207  
 Db 577 ----- 577  
 OY 1208 GCGACCTGGGCGCTTCCGCTGAGAGACGCGGTATGAAGATGACCGGCGCTGACCGCG 1267  
 Db 578 -----Pro 578  
 OY 1268 CGCGCTTCGAGCGCGCGCGCGCGCGCGCGCGACTGACGCGCGGTACTTCGCGACCTGTGG 1327  
 Db 579 LysAlaAlaAla---ProLysThrProLysGluProAlaProThrThrProLysGluPro 597  
 OY 1328 TGTTCGACCGCGCAACGAGGCGCGATGCG---CCACCTGCAACACCTTACCGAGCGG 1384  
 Db 598 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 617  
 OY 1385 CGCGCGCATCATTCGCTGTACGTCAACGCGCGCGCGTGTGGCAAGAGAGCGCTTCA 1444  
 Db 618 ProLysGluThrAlaPro---ThrThrLeuLysGluProAlaProThrThrProLysLys 636  
 OY 1445 CGCGCGACATGCGCGCGCGCTGCTGCACGACGCGCGCGCTGAGCGCGCGCGCGCT 1504  
 Db 637 ProAla-----ProLysGluLeuAlaProThrThrThrLysGluProThrSerThr--- 653  
 OY 1505 TACAATCGCGGTAGACGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1564

```

Db 654 -----|||::|||::|||
QY 1565 ACATGCGCCCTCCCTCCGTCGCAATACGACCCCGCATATGCGGAGAGATGA 1624
Db 669 ThrThrProLysGluProAlaProThrThrPro-----679
QY 1625 TGGCGCGCGCCGTCGCGGAGCGGCAAGCCCGGAAATGACCTGCAAGACCTGCGC 1684
Db 680 -----LysGluProAlaProThrThrProLysGluThrAlaPro---ThrThrLeuLys 696
QY 1685 AGCGCAGCG---GCATCGCGGCTGTCGACCTGTCACAG-----CCGAGCTGGGCC 1732
Db 697 GluProAlaProThrThrProLysLysProAlaProLysLysLysLysLysLysLys 716
QY 1733 AGATCCCGCCGACTACGAGAAC 1756
Db 717 LysGluProThrSerThrThrSer 724

RESULT 11
US-10-124-557-104
; Sequence 104 Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Casert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104
Alignment Scores:

```

```

Pred. No.: 7,66e-06 Length: 1140
Score: 285.50 Matches: 137
Percent Similarity: 33.50% Conservative: 60
Best Local Similarity: 23.30% Mismatches: 240
Query Match: 8.65% Indels: 151
Gaps: 24

US-10-009-782-1 (1-1758) x US-10-124-557-104 (1-1140)
QY 32 CCATGTCGCAATCCGATCCGACCCCTTCGACTGCTCGCGGCGGACCCGATCG 91
Db 366 ProThrProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 385
QY 92 ACAGCAGCA-----ACACCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 142
Db 386 ThrLysSerAlaProThrThrThrThrThrThrThrThrThrThrThrThrThr 400
QY 143 GCATCGCGCCGATCGCGGATCTGCGAGCGCGCGCGACACCGCGGTCGACGTGCG 202
Db 401 ThrLysGluProAlaPro-----ThrThrProLysGluProAlaProThr 415
QY 203 GCGTGGTGGTGGCGCGCGGCTTCATGACTCGACACCCACGACGAACTACTGCTCA 262
Db 416 -----ThrThrLysGluProAlaProThrThrThrThrThrThrThrThr 427
QY 263 GCGTGGCGCATGACGACCCGACGATCTCGAGGGCGTCAAGCGGTCAGCGGCAATT 322
Db 428 AlaProThrThr-----ProLysGluProAlaPro-----437
QY 323 GCGGCAATACGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 382
Db 438 -----ThrThrProLysLysLysLysLysLysLysLysLysLysLysLys 449
QY 383 ACAGAGCGGCTCTTACGCTTCGAGCGCTTCGCGACACCTGACGCGGTGCGGCA 442
Db 450 GluProAlaProThrThrThrThrThrThrThrThrThrThrThrThrThrThr 469
QY 443 CGCGGGCGG-----CGTCACGCGCGCGCTGATGTTGCGCATTCACGTCGCG 493
Db 470 ThrThrLysGluProAlaProThrThrPro-----479
QY 494 CGCGGTATGCGCGGACTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 553
Db 480 -----LysGluProAlaProThrThrAlaProLysProAlaPro 492
QY 554 ACCTGCGCGAGAGCATGCGCGGCGCGCATTCGACCGCGCTCTAC 613
Db 493 ThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThr 510
QY 614 CGCGCGCGCGCGCGCA---CCACCGAAGATCATGAGGTGCGCGCGCGCGCG 670
Db 511 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrThrLys 530
QY 671 CGCATGGCGGATCATGCGCACCCACATGCGCGAGAGGAGGACGACATCGTGGCGCG 730
Db 531 ProThrThrThrLysGluProAlaProThrThrThrThrThrThrThrThrThr 550
QY 731 TGGAGAAACCTTCGATCGCGCGCGCGAGCTGAGCGGTGATTCGACCA 790
Db 551 GluProSerProThr-----ThrThrLysGluProAlaProThrThr 564
QY 791 AGTCATGCGCGCGCAATTCGCGCGCTGCGCGAGACGCTCGCGTGTATGAGCGCG 850
Db 565 ProLysGluProAlaPro-----ThrThrPro 573
QY 851 CCATGCGCGCGCGAGAGCTGCTGCGAGCGGTATCCCTACGCGCGCGCGCGCATG 910
Db 574 LysLysProAlaProThrThrThrThrThrThrThrThrThrThrThrThrThr 593
QY 911 TCAAGCAGACCGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
Db 594 ProThrThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrThrPro 613

```

[illegible]

QY 443 CGCGGGG-----CCGTCAACGCCGCTGTATGTTGGCCATTCACAGCTGGCG 493  
 Db 336 ThrThrLysGluProAlaProThrThrPro----- 345  
 QY 494 CGCGGTCATCGCGGATTCAGCGCGCCAGCAGAGAAATCCGGCCATCGCG 553  
 Db 346 -----LysGluProAlaProThrThrProLysProAlaPro 358  
 QY 554 ACCTGCGCGAGAGAACCTAGCGCGCCGCGCATTCGCGCGCGCTTCTAC 613  
 Db 359 ThrThrLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 376  
 QY 614 CGCGCGCGCGCGCGCA---CAACCGAAGATCATGAGTGTGCGCGCGCTGACG 670  
 Db 377 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrLysSerAla 396  
 QY 671 CGCATGCGCGCATTCACCCACCATGCGAGAGCGGAGACATCGCGCGCGC 730  
 Db 397 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 416  
 QY 731 TGGAGGAACCTTCGCGATCGCGCGCGCGAGCTGACGTCGCGGTGATCGACACACA 790  
 Db 417 GluProSerProThr-----ThrThrLysGluProAlaProThrThr 430  
 QY 791 AGGTCAATGCGCGCAATTCGCGCGCTGCGCGAGACGCTGCGTATCGAGCGC 850  
 Db 431 ProLysGluProAlaPro----- 439  
 QY 851 CCATGCGCGCGCGCGCATCTGCTGAGACGGTATCCCTACGTGCGCGCTGACACATGC 910  
 Db 440 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 459  
 QY 911 TCAGAGGAGCGCGCTGCTGCTGCGCGCGAGCACATCATCAGCTGTGAGAGCGCTTCC 970  
 Db 460 ProThrThrThrLysLysProAlaProThrThrProLysGluProAlaProThrThrPro 479  
 QY 971 CCGAATGACGCGCGCGCGCATGATGAGTCGCGCGCGCGCGCGCGCAATCCAGATGC 1030  
 Db 480 LysGluThrAlaProThrThrProLysLysLeuThrProThrThrProGluLysLeuAla 499  
 QY 1031 ACCTGTCGCGCGAGCTGACGCGCGCGCGCGCATCTCATGATGAGAGACCGCGAGC 1090  
 Db 500 ProThrThrPro-----GluLysProAlaProThrThrProGluLysLeuAlaProThr 517  
 QY 1091 TGCAGGAGATCCAGGCTGCGCGCGCGCATGATGC---GCTCCGAGCGCGCTCGCGCAGC 1147  
 Db 518 ThrProGluLysProThrProThrThrProGluLysProAlaProThrThr----- 534  
 QY 1148 ACGAGCGCGCGCATCCGCGCTGTGGGACCTTCCGCGCGGTGTGGGCGCACTATGCGC 1207  
 Db 534 ----- 534  
 QY 1208 GCGACCTGGGCTGTCCCGCTGAGAGCGCGGTATGAAATGACCGCGCTGACCGCGC 1267  
 Db 535 -----Pro 535  
 QY 1268 CGCGCTGCGCGCTGCGCGCGCGCGCGCACTGACGCGCGGTACTTCCCGACCTGTGTG 1327  
 Db 536 LysAlaAlaAla---ProsmThrProLysGluProAlaProThrThrProLysGluPro 554  
 QY 1328 TGTTCGACCGCGCGAGGCGCGATACG---CCACCTTCGACACCTACCGAGCGG 1384  
 Db 555 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 574  
 QY 1385 CGCGCGCGCATTCCTCGGTACGTCACAGCGCGCGCGCTGTGGAAGAGCGAGCGTTCA 1444  
 Db 575 ProLysGluThrAlaPro---ThrThrLeuLysGluProAlaProThrThrProLysLys 593  
 QY 1445 CGCGCGAGCATCGCGCGCGCTGCGAGCGAGCGCGCGCGCTGAGCGCGCGCGCGCGCT 1504  
 Db 594 ProAla-----ProLysGluLeuAlaProThrThrThrLysGluProThrSerThr--- 610

QY 1505 TACAATCCGGGTTAAGGGGGGCGGTCCGCCCTCCCAACCTGAGCGCAACCGCT 1564  
 Db 611 -----ThrSerAspLysProAlaProThrThrProLysGluThrAlaPro 625  
 QY 1565 ACATGCGCGCTCCCTCCCTGCGCATACGCGCGCGCGCATATGCTGGGCAAGAAATGA 1624  
 Db 626 ThrThrProLysGluProAlaProThrThrPro----- 636  
 QY 1625 TGGGGCGCGCGCTCGCGCGCGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1684  
 Db 637 -----LysGluProAlaProThrThrProLysGluThrAlaPro---ThrThrLeuLys 653  
 QY 1685 AGCGACGCG---GCATGCGCGCTGACCGCTGCGCAAG-----CGAGCTGGCGC 1732  
 Db 654 GluProAlaProThrThrProLysLysProAlaProLysGluLeuAlaProThrThrThr 673  
 QY 1733 AGATGCGCGCTGAGCTACGAGAGC 1756  
 Db 674 LysGluProThrSerThrThrSer 681

# RESULT 13

US-10-124-557-42  
 Sequence 42, Application US/10124557  
 Patent No. US20020137894A1

## GENERAL INFORMATION:

APPLICANT: Turner, Katherine  
 Clark, Stephen C.  
 Jacobs, Kenneth  
 Hewick, Rodney G.  
 Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 NUMBER OF SEQUENCES: 143  
 CORRESPONDENCE ADDRESS:

ADDRESSER: Genetics Institute, Inc.  
 STREET: 87 Cambridgepark Drive  
 CITY: Cambridge  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02140

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557  
 FILING DATE: 16-Apr-2002  
 CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502  
 FILING DATE: 18-JAN-1991  
 APPLICATION NUMBER: US 07/546,114  
 FILING DATE: 29-JUN-1990  
 APPLICATION NUMBER: US 07/457,196  
 FILING DATE: 29-DEC-1989  
 APPLICATION NUMBER: US 07/390,901  
 FILING DATE: 08-AUG-1989

## ATTORNEY/AGENT INFORMATION:

NAME: Cseer, Luann  
 REGISTRATION NUMBER: 31,822  
 REFERENCE/DOCKET NUMBER: GI 5190  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)876-1170  
 TELEFAX: (617)876-5851  
 INFORMATION FOR SEQ ID NO: 42:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1311 amino acids  
 TYPE: amino acid

MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 42:  
 US-10-124-557-42

## Alignment Scores:

Pred. No.: 7.56e-06 Length: 1311  
 Score: 285.50 Matches: 137  
 Percent Similarity: 33.50% Conservative: 60  
 Best Local Similarity: 23.30% Mismatches: 240  
 Query Match: 8.65% Indels: 151  
 DB: 12 Gaps: 24

US-10-009-782-1 (1-1758) x US-10-124-557-42 (1-1311)

QY 32 CCATGTCACCAATCCGATCCAGCCCTTCGACCTGCTGCTCCGCGGCGCACCTTCATCG 91  
 DB 273 ProthThrProthThrThrLysSerAlaProthThrProLysGluProAlaProthThr 292  
 QY 92 AGGCGACA-----ACACCCGCGGCGCGCGCGCACCTGGGCGCGCGCGCAC 142  
 DB 293 ThrLysSerAlaProthThrProLysGluProAlaProthThr-----Thr 307  
 QY 143 GCATGCGCGCCATCGGCATCTGTGAGACCGCGCGCGCACACCGGGTGCACGTGTGCG 202  
 DB 308 ThrLysGluProAlaPro-----ThrThrProLysGluProAlaProthThr----- 322  
 QY 203 GCGTGTGATGCGCGCGCTTCATCGACTGCGACACCGCACAGACACACTACTGCTCA 262  
 DB 323 -----ThrThrLysGluProAlaProthThrThrLysSer 334  
 QY 263 GCGCGCGACATAGCCCAAGATCTGCGAGGGCGTCCACCGGTGTGTCAGGGCAATT 322  
 DB 335 AlaProthThr-----ProLysGluProAlaPro----- 344  
 QY 323 GCGGCATGACCTGTGGCGCGCTGGCGACCGCCACCGCGCGCGCGCGCGCTGTGCG 382  
 DB 345 -----ThrThrProLysGluProAlaProthThrProLys 356  
 QY 383 ACGAAGGCGCTTACCGCTTTCAGCGCTTCGCGACCTGACCTGAGCGCTGTGGGCGCA 442  
 DB 357 GluProAlaProthThrProLysGluProthProthThrProLysGluProAlaPro 376  
 QY 443 CGCGCGCG-----CGTCAACCGCGCGCTGTATGAGGGCCATTCAACGCTGCGCG 493  
 DB 377 ThrThrLysGluProAlaProthThrPro----- 386  
 QY 494 CCGCGGTATGCGGACTTCAGCGCGCGCGCACCGACGAGAAATCGCGCGCATGCGGG 553  
 DB 387 -----LysGluProAlaProthThrAlaProLysGluProAlaPro 399  
 QY 554 ACCTGCGCGAGAGACGATGCGAGCGCGCGCATTCGACCGCGCGCTTCTACG 613  
 DB 400 ThrThrProLysGluProAlaProthThrProLys-----GluProAlaProthThr 417  
 QY 614 CGCGCGCGCGCGCA---CGACGAGAGATCATCGAGGTGCGCGCGCGCTGAGCG 670  
 DB 418 ThrLysGluProSerProthThrProLysGluProAlaProthThrThrLysSerAla 437  
 QY 671 CGCATGCGGACATCTAGCCACCCACATCGCGAGAGCGGAGCGACATCGGCGCGCG 730  
 DB 438 ProthThrThrLysGluProAlaProthThrThrLysSerAlaProthThrProLys 457  
 QY 731 TGGAGAAACCTTCGCGATCGCGCGCGAGCTGAGTGTGATCTGCGACACA 790  
 DB 458 GluProSerProthThr-----ThrThrLysGluProAlaProthThr 471  
 QY 791 AGGTATGCGCGCAATTCGCGCGCTGCGCGAGACGCTGCTGATCGAGGCGG 850  
 DB 472 ProLysGluProAlaPro-----ThrThrPro 480  
 QY 851 CCATGCGCGCGCGACGATCTGCTGAGCGGTATCCCTAGTGCGCGCGCTCCACACG 910  
 DB 481 LysLysProAlaProthThrProLysGluProAlaProthThrProLysGluProAla 500  
 QY 911 TCAACAGAGAGCGCTGCTGCGCGCGACCGACCATCATCACTGTGTGACCGCTTCC 970  
 DB 501 ProthThrThrLysLysProAlaProthThrAlaProLysGluProAlaProthThrPro 520

QY 971 CGGAACGTAGCGGCGGACCTGTGATGAATGCGGCGCGAGCGGCAATTCAGTACG 1030  
 DB 521 LysGluThrAlaProthThrThrProLysLysLeuThrProthThrProLysLeuAla 540  
 QY 1031 ACGTGTGCGCGAGCTGACGCGCGCGCGCGCATCTACTGATGATGACGACCGGAG 1090  
 DB 541 ProthThrPro-----GluLysProAlaProthThrProLysGluLeuAlaProthThr 558  
 QY 1091 TGCACGGATCTGTGCGCTTCCGCGCGCGACCATGATCG---GCTCGAGCGCGCTGCGCAG 1147  
 DB 559 ThrProGluGluProthThrProthThrProLysGluProAlaProthThr----- 575  
 QY 1148 ACGAGCGCGCGCATCGCGCGCTGTGGGCGACCTTCCGCGGCGTGTGCGGACATAAGCGC 1207  
 DB 575 ----- 575  
 QY 1208 GCGACCTGCGCGCTTCCGCGTGAAGACGCGGATGAGATGACCGCGCTGACCGCG 1267  
 DB 576 -----Pro 576  
 QY 1268 GCGCGTTCGCGCTGCGCGCGCGCGGCGGACGTGACAGCGGATTCGCGGACCTGTGCG 1327  
 DB 577 LysAlaLeuAla---ProAsnThrProLysGluProAlaProthThrProLysGluPro 595  
 QY 1328 TGTTCGACCGCGCGCGCGCGCGATACCG---CCACCTTCGAAACACCTACGAGCGCG 1384  
 DB 596 AlaProthThrProLysGluProAlaProthThrProLysGluThrAlaProthThr 615  
 QY 1385 CCGCGCGCACCATTCGCTTACGTCAACGCGCGCGCGGTGCGACAGACGAGCGCTTCA 1444  
 DB 616 ProLysGlyThrAlaPro---ThrThrLeuLysGluProAlaProthThrThrProLys 634  
 QY 1445 CCGCGACGATGCGCGCGCGCGCGTGTGCGACGAGCGCGCGCGCGCGCGCGCGCG 1504  
 DB 635 ProAla-----ProLysGluLeuAlaProthThrThrLysGluProthThrSer 651  
 QY 1505 TCAATCCGCGCGGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1564  
 DB 652 -----TherAspLysProAlaProthThrThrProLysGlyThrAlaPro 666  
 QY 1565 ACATGCG 1624  
 DB 667 ThrThrProLysGluProAlaProthThrPro----- 677  
 QY 1625 TGGCG 1684  
 DB 678 -----LysGluProAlaProthThrThrProLysGlyThrAlaPro---ThrThrLeuLys 694  
 QY 1685 AGCGCAGCG---GCATCGCGGCTTCGACCGCTGTCCAGG-----CCGAGCTGGCGC 1732  
 DB 695 GluProAlaProthThrThrProLysLysProAlaProLysGluLeuAlaProthThrThr 714  
 QY 1733 AGATGCG 1756  
 DB 715 LysGlyProthThrSerThrThrSer 722

# RESULT 14 US-10-124-557-142

Sequence 142, Application US/10124557  
 Patent No. US20020137894A1

## GENERAL INFORMATION:

APPLICANT: Turner, Katherine  
 Clark, Stephen C.  
 Jacobs, Kenneth  
 Hewick, Rodney M.  
 Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 NUMBER OF SEQUENCES: 143  
 CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.  
 STREET: 87 Cambridgepark Drive  
 CITY: Cambridge



```

STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 1313 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142

Alignment Scores:
Pred. No.: 7.56e-06 Length: 1313
Score: 285.50 Matches: 137
Percent Similarity: 33.50% Conservative: 60
Best Local Similarity: 23.30% Mismatches: 240
Query Match: 8.65% Indels: 151
Gaps: 24
DB: 12

US-10-009-782-1 (1-1758) x US-10-124-557-142 (1-1313)
QY 32 CGATGCCCAATCCGATGCCAGCCCTTGACCTGCTGCGGGCGGCACCCCTCATCG 91
DB 275 ProthProthThrThrLysSerAlaProthThrThrProLysGluProAlaProthThr 294
QY 92 ACGGAGCA-----ACACCCGGGGGGGGCGCGGACCACTGGGGCGTGGCGGAC 142
DB 295 ThrLysSerAlaProthThrThrProLysGluProAlaProthThr-----Thr 309
QY 143 GCATGCCCGCATCGCGATCTGTGGACGCGCGCGACACACCGCGGTGACGTGTCGG 202
DB 310 ThrLysGluProAlaPro-----ThrThrProLysGluProAlaProthThr----- 324
QY 203 GCGTGTGGTGGCGCGCGCTCATCGATCGACACCAACAGCAACAATACTGCTGCA 262
DB 325 -----ThrThrLysGluProAlaProthThrThrThrLysSer 336
QY 263 GCGCGTGGACATGAGCCGCAAGATCTGCGAGGGGGTACACGGTGGTACAGGGCAATT 322
DB 337 AlaProthThr-----ProLysGluProAlaPro----- 346
QY 323 GCGGCATACGCTGGCGCGCTGGCGACGCCACCCCGCGCCCTGAGACTGCTGG 382
DB 347 -----ThrThrProLysLysProAlaProthThrThrProLys 358
QY 383 ACGAAGGCGGCTTACCGTTTCGAGCGCTTGCGGACTGACGCGGTGGCGGCCA 442
DB 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

```

```

DB 359 GluProAlaProthThrThrProLysGluProthThrProthThrProLysGluProAlaPro 378
QY 443 CGCCGCGG-----CGGTCAACGGCGGCTGATGTGGGCATTCAACGCTGGCG 493
DB 379 ThrThrLysGluProAlaProthThrPro----- 388
QY 494 CCGCGGTATCCGGAATGTGACGGCGCGCCACCAAGACAGAGAAATGCGGCATGGCGG 553
DB 389 -----LysGluProAlaProthThrAlaProLysLysProAlaPro 401
QY 554 ACCTGGCCGAGAAACCATGGCCACAGCGGCCCATGGCATTTGACCGGCGCTTCTAC 613
DB 402 ThrThrProLysGluProAlaProthThrThrProLys-----GluProAlaProthThr 419
QY 614 GCGCCGCGCGCGGCGCA---CAACGAGAGATCATCGAGTGTGCGCGCGCTGAGCG 670
DB 420 ThrLysGluProSerProthThrThrProLysGluProAlaProthThrThrLysSerAla 439
QY 671 CGCATGGCGCATCTACGCCACCCACATGCGCGAGAGAGGCGACATGCGCGCGCG 730
DB 440 ProthThrThrLysGluProAlaProthThrThrLysSerAlaProthThrProLys 459
QY 731 TGGAGAAACCTTCCGATGCGCGCGCGGACGTCGCGGTGATTCGACACACA 790
DB 460 GluProSerProth-----ThrThrLysGluProAlaProthThr 473
QY 791 AGGTATGGCGACCCCAATTTGCGCGCGTGGCGGAGAGAGGTGCGGCTGATGAGCGCG 850
DB 474 ProLysGluProAlaPro-----ThrThrPro 482
QY 851 CCATGGCGCGCCAGAGAGCTGTGCGTGACGATCCATGTCGCGCGGTCCACACATGC 910
DB 483 LysLysProAlaProthThrThrProLysGluProAlaProthThrProLysGluProAla 502
QY 911 TCAAGCAGACCGGCTCTCTGCGCGGACGACATCATCAGCTGTGTCGACCCCTTC 970
DB 503 ProthThrThrLysLysProAlaProthThrAlaProLysGluProAlaProthThrPro 522
QY 971 CCGAATGACGGCGCGCGGACCTGTGATGATCGCGCGGCGGCGCAATTCAGATGAC 1030
DB 523 LysGluThrAlaProthThrThrProLysLysLeuThrProthThrProLysLysLeuAla 542
QY 1031 ACGTGGCGCGGACGTGACGCGCGCGCGGCGCATCTACTCATGATGAGAGAACCGGAG 1090
DB 543 ProthThrPro-----GluLysProAlaProthThrThrProLysLysLeuAlaProth 560
QY 1091 TGCAGGCATCTGCGGCTGCGCGCGGACATGATG---GTCGCGAGCGGCTCGGACAG 1147
DB 561 ThrProGluThrProthThrProthThrProGluThrProAlaProthThr----- 577
QY 1148 ACGAGCGCGCGCATCCGCGCTGTGGGACATTCGCGGCGTGGGGCACTATGCGC 1207
DB 577 ----- 577
QY 1208 GCGACTGGGCGCTGCCGCTGGAGACGCGGATGAGAAATGACCGGCTGACGCGCG 1267
DB 578 -----Pro 578
QY 1268 CGCGTTCGCGCTGCGCGCGGCGGCGACGTGACAGCGCGGATTCGCGGACCTGGTGG 1327
DB 579 LysAlaAlaAla---ProsnThrProLysGluProAlaProthThrProLysGluPro 597
QY 1328 TGTTCAGCCCGGCGACGATGCGCATACG---CCACCTTCAGAACCTTACGAGCGG 1384
DB 598 AlaProthThrThrProLysGluProAlaProthThrThrProLysGluThrAlaProthThr 617
QY 1385 GCGCGGCGATTCATTCGCTACGTCAACAGGCGCGCGCTGCGCAAGACAGCGTTCA 1444
DB 618 ProLysGluThrAlaPro---ThrThrLeuLysGluProAlaProthThrThrProLysLys 636
QY 1445 CCGGCGAGCATGCGCGCGGTGTCGACGACGAGCGCGCGGCGGCGGCGGCGGCGCT 1504
DB 637 ProAla-----ProLysGluLeuAlaProthThrThrLysGluProthThrSerThr--- 653

```

LEUCININ: 1314 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein.  
 SEQUENCE DESCRIPTION: SEQ ID NO: 50  
 US-10-124-557-50

Alignment scores:	
Pred. No.:	7,56e-06
Score:	265,50
Percent Similarity:	33,50%
Best local Similarity:	23,30%
Query Match:	8,65%
DB:	12
	5
	Gaps:
	24

OS-10-009-82-1 (1-1758) X OS-10-124-557-50 (1-1314)

[illegible]



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: May 11, 2003, 12:10:05 ; Search time 36.5 Seconds

(without alignments)  
2834.274 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 3299  
Sequence: 1 gaattccactgacgcgga.....ccctgagctacgagaagcct 1758

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 26574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2\_1/USP00/US10009782/runat\_07052003\_122518\_23179/app\_query.fasta\_1.1927  
-DB=Issued\_Patents\_AA -OEXT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=Dits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdl  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10009782 -ECGN 1.1.21 -runat\_07052003\_122518\_23179 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGEDQUERY -NEG\_SCORES=0 -WAIT -LONELOG -DEV\_TIMEROUT=120  
-WARN\_TIMEROUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued Patents AA:

1: /cgn2\_6/ptodata/1/aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/aa/PCRTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	398.5	12.1	1213 4 US-09-413-814-79	Sequence 79, Appl
2	344	10.4	882 4 US-09-413-814-78	Sequence 78, Appl
3	320.5	9.9	882 4 US-09-413-814-78	Sequence 78, Appl
4	310	9.4	558 4 US-09-199-637A-277	Sequence 277, Appl
5	309	9.5	1213 4 US-09-413-814-79	Sequence 79, Appl
6	287.5	8.7	960 4 US-09-219-849-6	Sequence 6, Appl
7	285.5	8.7	941 4 US-07-757-022B-14	Sequence 14, Appl
8	285.5	8.7	1022 4 US-07-757-022B-84	Sequence 84, Appl
9	285.5	8.7	1038 4 US-07-757-022B-74	Sequence 74, Appl
10	285.5	8.7	1049 4 US-07-757-022B-58	Sequence 58, Appl
11	285.5	8.7	1140 4 US-07-757-022B-104	Sequence 104, Appl
12	285.5	8.7	1270 4 US-07-757-022B-44	Sequence 44, Appl

13	285.5	8.7	1311 4 US-07-757-022B-42	Sequence 42, Appl
14	285.5	8.7	1313 4 US-07-757-022B-142	Sequence 142, Appl
15	285.5	8.7	1314 4 US-07-757-022B-50	Sequence 50, Appl
16	285.5	8.7	1320 4 US-07-757-022B-46	Sequence 46, Appl
17	285.5	8.7	1320 4 US-07-757-022B-60	Sequence 60, Appl
18	285.5	8.7	1334 4 US-07-757-022B-48	Sequence 48, Appl
19	285.5	8.7	1361 4 US-07-757-022B-40	Sequence 40, Appl
20	285.5	8.7	1363 4 US-07-757-022B-52	Sequence 52, Appl
21	285.5	8.7	1404 4 US-07-757-022B-2	Sequence 2, Appl
22	285.5	8.7	1404 4 US-07-757-022B-62	Sequence 62, Appl
23	283	8.6	1065 1 US-08-642-255-72	Sequence 72, Appl
24	277	8.4	960 4 US-09-219-849-5	Sequence 5, Appl
25	276.5	8.4	1057 3 US-08-831-820-1	Sequence 1, Appl
26	275	8.3	1064 1 US-08-642-255-62	Sequence 1, Appl
27	270	8.2	1461 4 US-09-585-887-9	Sequence 9, Appl
28	270	8.2	1461 4 US-09-289-578-9	Sequence 9, Appl
29	267.5	8.2	1958 1 US-07-945-283-2	Sequence 2, Appl
30	267	8.1	720 4 US-09-219-849-4	Sequence 4, Appl
31	267	8.1	777 1 US-08-642-255-53	Sequence 53, Appl
32	267	8.1	1341 3 US-08-663-825-18	Sequence 18, Appl
33	267	8.1	1341 4 US-09-500-811-18	Sequence 18, Appl
34	267	8.1	1341 4 US-09-570-573-18	Sequence 18, Appl
35	267	8.1	1341 4 US-09-548-608-18	Sequence 18, Appl
36	266	8.1	552 4 US-09-219-849-7	Sequence 7, Appl
37	263	8.0	822 4 US-09-219-849-49	Sequence 49, Appl
38	262.5	8.0	633 1 US-08-642-255-73	Sequence 73, Appl
39	262	7.9	1958 1 US-07-945-283-2	Sequence 2, Appl
40	257.5	7.9	355 4 US-08-483-533-41	Sequence 41, Appl
41	257.5	7.9	355 4 US-09-283-471A-41	Sequence 41, Appl
42	257.5	7.9	355 5 PCT-US91-06532-3	Sequence 3, Appl
43	254.5	7.7	395 4 US-09-370-838-187	Sequence 187, Appl
44	252	7.6	1442 2 US-08-316-650-12	Sequence 12, Appl
45	252	7.6	1442 5 PCT-US95-02251-12	Sequence 12, Appl

#### ALIGNMENTS

RESULT 1  
US-09-413-814-79  
Sequence 79, Application US/09413814  
Patent No. 6225064  
GENERAL INFORMATION:  
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
APPLICANT: Bristol-Myers Squibb, Co.  
APPLICANT: Beyer, Stefan  
APPLICANT: Bloeker, Helmut  
APPLICANT: Brandt, Petra  
APPLICANT: Cino, Paul M  
APPLICANT: Dougherty, Brian A  
APPLICANT: Goldberg, Steven L  
APPLICANT: Hoffe, Gerhard  
APPLICANT: Mueller, Joachim  
APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or  
FILE REFERENCE: PCT/US 99/23535  
CURRENT APPLICATION NUMBER: US/09/413,814  
EARLIER FILING DATE: 1999-10-07  
EARLIER APPLICATION NUMBER: DE 198 46 493.2  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 79  
LENGTH: 1213  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-413-814-79  
Alignment Scores:  
Pred. No.: 1,416-19  
Score: 398.50  
Percent Similarity: 36.428  
Best Local Similarity: 32.258  
Length: 1213  
Matches: 209  
Conservative: 27  
Mismatch: 230

```

Query Match::      12.08%      Indexs:      182
DB:                4          Gaps:           36

US-10-009-782-1 (1-1758) x US-09-413-814-79 (1-1213)

```

[illegible][illegible]

RESULT 2  
US-09-413-814-78  
; Sequence 78, Application US/09413814  
; Patent No. 625064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Bayer, Stefan  
; APPLICANT: Bloecker, Helmut









```

OY 879 GCGTATCCCTACGTGGCGGCTCCACCATGCTCAAGACAGACCGGCTGCTGCGCGG 938
DB 322 GYGLIUGLIPROSERGLCYSLIEGLNHISERPROGLYALAAG-----GLYArg 339
OY 939 ACAGACCATCATACCTCGGTGCAAGCCCTCCCGAATGAGGCGGCGACCTGATGA 998
DB 340 ThrPro-----AspArgLInSerArgLArgLArgLLeu 351
OY 999 AGTCGGCGGCGCA---GCGCGGCAATCCAGTACGAGCTGGTCCGAGCTGACGCGG 1055
DB 352 ProGLNHISArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 371
OY 1056 GCGCGG---CATCTACTTCATGATGAGACGAGAACCGGACGAGTGCATGCT 1103
DB 372 ArgArgProArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 391
OY 1104 GCGGT---CGGCGCGACCAT-----GAT 1124
DB 392 GYLIIEHIEGLNProAspArgLArgLArgLArgLArgLArgLArgLArgLArgL 411
OY 1125 GCGGTCCGACGCGCTGCGCGACGAGACGCGCGCGCATCGCGCTGGGCGACCTTCC 1184
DB 412 ArgArgHISArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 426
OY 1185 GCGGTGCTGCGGACATGCGCGGACCTGCGGCTGCTCCGCTGAGACGCGGTATG 1244
DB 427 -----GlyArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 441
OY 1245 GAGATGACCGGCTGACCGCGCGCGCTTGGCGCTGC-----CGGCGCGCGGCACT 1298
DB 442 -----GlnAlaAspCysArgLArgLArgLArgLArgLArgLArgLArgLArgL 458
OY 1299 GCGACCGCG-----GTACTTCGCGGCACTGTGTGTGCA---CGGCGCGACGCGTGC 1349
DB 459 AlaArgLISArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 475
OY 1350 CGATACCGCACCTTCGACACCTACCGA-----GCGCGCGCGCGCAT 1394
DB 476 -----ArgLInLIEArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 493
OY 1395 CCAATCCGCT-ACGTCAACGCGCGCGCGCTGCGCAAGACGCGGCTTACCGCGCAC 1453
DB 494 ArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArg 509
OY 1454 ATCCGCGCGCGTGCACGACGCGCGCGCTGACCGCGCGCGCGCTTACATCGG 1513
DB 510 -----AlaGLYLeuPheHISGLYArgPro----- 517
OY 1514 GCGTGAAAGCGGCGCGCTGCGCGCGCGCTCCCAACCTGAGCGCAACCGCTACATGCC 1573
DB 518 -----ValGLYAlaArgLArgL----- 522
OY 1574 CTCCCTCCGCTCGCAATACGCGCGCGCATATGCTGGGCAAGGAG 1621
DB 533 LeuProHISAlaValArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 538

```

## RESULT 5

US-09-413-814-79

Sequence 79 Application US/09413814

Patent No. 6223064

GENERAL INFORMATION:

APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH

APPLICANT: Bristol-Myers Squibb, Co.

APPLICANT: Beyer, Stefan

APPLICANT: Bloeker, Helmut

APPLICANT: Brandt, Petra

APPLICANT: Cino, Paul M

APPLICANT: Dougherty, Brian A

APPLICANT: Goldberg, Steven L

APPLICANT: Hofler, Gerhard

APPLICANT: Mueller, Joachim

APPLICANT: Reichenbach, Hans

```

; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 1213
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-413-814-79

```

## Alignment Scores:

```

Pred. No.: 1,73e-13 Length: 1213
Score: 309.00 Matches: 204
Percent Similarity: 32.13% Conservative: 9
Best Local Similarity: 30.77% Mismatches: 250
Query Match: 9.53% Indels: 200
DB: 4 Gaps: 35

```

US-10-009-782-1 (1-1758) x US-09-413-814-79 (1-1213)

```

OY 1756 GCTTCGTAGCTACAGGCGCATCTGCGCGGCTGCGCGGCTGCGCGGCGG 1697
DB 18 AlaArgArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArg 37
OY 1696 TCGCGTGCCTGCGACAGGCTTGCAGGCTTCCGCGCTGCGCGGCGCA 1637
DB 38 ArgArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 55
OY 1636 GCGCGCGCGCGCATCTCTTCCGCGCGGCGGCGGCGGCGGCGGCGGCGG 1580
DB 56 AlaArgArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArg 71
OY 1579 -----AGGAGGCGCGCATGTAGCGGCTTGCAGGCGGTTG 1544
DB 72 AlaProGLNAlaProProGLNAlaProGLNAlaProGLNAlaProGLNAla 88
OY 1543 GAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1484
DB 89 ValGLYGLY-----ArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 102
OY 1483 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1433
DB 103 AlaPro-----ArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 121
OY 1432 CTGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1388
DB 122 -----AlaArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 137
OY 1387 -----CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1358
DB 138 ValHISGLYAlaArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 157
OY 1357 CGGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1298
DB 158 ArgHISGLNProProLeu-----ArgLArgLArgLArgLArgLArgLArgLArg 175
OY 1297 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1253
DB 176 ValAlaGLNLeuArgProValArgLArgLArgLArgLArgLArgLArgLArg 195
OY 1252 TCATCTTCATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1229
DB 196 ArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgL 215
OY 1228 -----GCGGGA-----ACAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1196
DB 216 AlaArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArgLArg 235

```

```

QY 1196 ----- 1196
Db 236 ValProArgArgGlyArgProGlyAlaValAlaProProArgHisAspGluArg 255
QY 1195 -----CCAGCACCAGCGGGAAGTGC---CCACAGAGCGCGGATGCGGCGTCTGCT 1145
Db 256 LeuAspProAlaGluAlaProArgSerHisProArgGlyLeuProAlaProArg 275
QY 1144 GCGGACAGCGGTGGAGCCATCATGTCG-----GCGCGA 1109
Db 276 GluProArgArgArgAspArgGlyValAspLeuValAspArgProProAspArgGlyArg 295
QY 1108 ACAGCAGAGATGCTGCTGAGCATCGGCTGCATCATGAGTACATGCGCGCGCT 1049
Db 296 ArgProGlyValAlaGlnHisProLeuArgProPro-----AlaArgGluProAla 312
QY 1048 GCAGCTGCGGCACACGCTCTACTTGATTTGCGCGCTGCG---CCGCGACTCATCA 992
Db 313 HisLeuArgAlaArgArgGlyAlaGlyAlaValSerAspArgGlyProArgArgAspPro 332
QY 991 GGTGCGCGCGCTCAGTTGGGGAGGCGTTGCACACAGTGATGATGTCGTCGCGCA 932
Db 333 HisArgArgArgArgArgArgArgAlaArgGlyLeuAlaArg----- 345
QY 931 GCAGCAGCGGCTCTGAGCATGGTGGAGCGCGCACGTAGGATACGCGCTCAGCG 872
Db 346 -----ArgGlyAlaAspProGlyAlaValProGluAlaProHisAspArgAlaAla 363
QY 871 -----AGAGCTCTGCGCGCGCGCATGCGCGCTGCATACGCGCGCTCGCGCG 821
Db 364 ValGluAspArgArgProGlyProLeuLeuArgArgArgHisAspArgAlaAlaAla 383
QY 820 AGCGCGCAATGGGCTGCGCATGACCTTGTGGCGAGATACACCGCGACGTCCA 761
Db 384 HisArg-----ProSerGly-----GluAspProArgGluPro 394
QY 760 GCTCGCGCGCGCATGCGGAAGTTTCTCCAGCGCGCGCACAGATGCTGCTGCTGCG 701
Db 393 HisArgAlaArgArgArgArg-----GlyArgProArgAlaAlaProLeuGlyArg 411
QY 700 GCATGCGGATGGCGTATGATCCGCCATGCGCGCTCAGCGCGCGCACCTCGATATCT 641
Db 412 AlaGlyGly-----ArgGlyGlyGluAspArgSerVal----- 422
QY 640 CTTCGGTGTGGCGCGCGCGCGCG---GGTAGA---AGCGCGCGGTGGAATGCCGA 587
Db 423 -----GlyArgGluAlaProGlyArgValAlaArgArgArgArgArgArg 437
QY 586 TGGCGCGCGCTGCGCATGCGCTTCTCGCGCGCGCATGCGCGCGCATGCTCTGCTGCG 527
Db 438 TTPArgArgAlaProArgLeuArgProGluGluAlaGlyValHisAspProArgGly 457
QY 526 TGGCGCGCGGTGCAAGTCCGCGCATGCGCGCGCGCGCGCGCTGTGAATGCCCATAC 467
Db 458 GlyArgArgProProGlyAlaAlaAlaGluArgGluArg-----GlnGlyGlyProArg 475
QY 466 AGG-----CGCGCTTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 437
Db 476 ArgProAlaGlySerArgArgGlyArgAlaAlaAlaArgArgGlyArgArgProAla---Asp 494
QY 436 GCAGCGCTTCCAGAGTACGCGCGGAAGCGCTCGAAGCGT----- 398
Db 495 ArgAspArgAlaAlaHisArgLeuGlyAlaArgGlyGlyLeuAlaGlyArgGlyArg 514
QY 397 -----AAGAGCGCGCTTCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 356
Db 515 ArgAspArgGluProLeuArgAlaArgLeuHisLeuAlaAlaAlaProArgAlaAlaAla 534
QY 355 TGGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296
Db 534 ----- 534
QY 295 CCGTGGAGATCTTGGGCGTCAATGCGGACCGCTGAGCAGGATGTTGTCGTCGTCGTC 236

```

```

Db 535 ProAlaArg-----ArgAlaHisArgGlyAlaAlaArgAlaArgArgGlyAla 552
QY 235 GCGAGTGATGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 176
Db 553 GlyArgValAlaAspArgProValProValProAspHisArgAlaHisGlyAlaAla 572
QY 175 CGCGCTCCGACAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 116
Db 573 ArgArgGlyLysArgGlyGln-----GlyArgAlaGlyArgArgArgGlyAlaAlaArg 590
QY 115 GCGCGCGCGCG 107
Db 591 GlyAlaPro 593

RESULT 6
US-09-219-849-6
/ Sequence 6, Application US/09219849
/ Patent No. 615081
/ GENERAL INFORMATION:
/ APPLICANT: VAN HEERDE, GEORGE V.
/ APPLICANT: VAN RIJN, ALEXIS C.
/ APPLICANT: BOWSTRA, JAN B.
/ APPLICANT: DE WOLF, FREDERIK A.
/ APPLICANT: MOEBROEK, ANDREAS
/ APPLICANT: WIND, RICHEL D.
/ APPLICANT: VAN DEN BOSCH, TANJA J.
/ TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
/ TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
/ FILE REFERENCE: PREPARATION THEREOF
/ CURRENT APPLICATION NUMBER: US/09/219,849
/ CURRENT FILING DATE: 1998-12-23
/ NUMBER OF SEQ ID NOS: 50
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 6
/ LENGTH: 960
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Illustrative
/ OTHER INFORMATION: amino acid sequence
US-09-219-849-6

Alignment Scores:
Pred. No.: 4,74e-12 Length: 960
Score: 287.50 Matches: 196
Percent Similarity: 35.45% Conservative: 33
Best Local Similarity: 30.34% Mismatches: 262
Query Match: 8.71% Indels: 157
DB: 4 Gaps: 40

US-10-009-782-1 (1-1758) x US-09-219-849-6 (1-960)
QY 63 CCGTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122
Db 205 ProGlyAlaGlnGly-----ProAlaGlyProGlyGlySerArgAsp 218
QY 123 CCGTGGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
Db 219 ProGly-----ProProGlyAlaHis---GlyProAlaGlyProGlyGlyAla 233
QY 183 CAC-----CCGCGTGCAGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 230
Db 234 HisGlyProAlaGlyProGlyGlyAlaHisGlyProAlaGlyProGlyGlyAlaHisGly 253
QY 231 CTCGACACCGCAGCA---CGCAACTACTGCT---CAGCGTGCAGCATGACGCCCA 284
Db 254 ProAlaGlyProGlyGlyAlaGlnGlyProAlaGlyProGlyGlySerArgAspProGly 273
QY 285 GATCTCGAGGCGGTGACACAGGTGTGTCAAGGCAATTGGCGCATGACGCTGCGCGCGCT 344

```

```

Db 274 ProProglYAlaInglYProAlaGlYProGlYglYserArg---AsPProGlYProPro 292
QY 345 GGGGACGCGCAACCGCC----- 362
Db 293 GlYAlaInglYProAlaInglYProGlYglYserArgAsPProGlYProAlaIn 312
QY 363 CCCCCCTGGAGCTGTGTGAGCAGAGCGGCTTACCGCTTGCAGCGCTTCCGCGACTA 422
Db 313 GlYProAlaInglYProGlYglYserArgAsPProGlYPro----- 325
QY 423 CCTGGAGCG-----GTTCGGGGCGCGCGCGCGCGCTGACAGCGCGCTGTATGGGG 476
Db 326 ProGlYAlaInglYProAlaInglYProGlYglYserArgAsP-----ProGlY 341
QY 477 CATTTCACGCTGCGCGCGCGCTGACAGCGCGCTGACAGCGCGCGCGCGCGCGCG 530
Db 342 ProProGlYAla-----GlnGlYProAlaInglYProGlYglYserArgAsPProGlYPro 359
QY 531 CGAGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 589
Db 360 ProGlYAlaInglYProAlaInglYProGlYglYserArgAsPProGlYPro 378
QY 590 GCATTTCGACCGCGCG-----CCTTCTACCGCGCGCG 619
Db 379 AlaInglYProAlaInglYProGlYglYserArgAsPProGlYProProGlYAlaInglY 398
QY 620 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 673
Db 399 ProAlaInglYProGlYglYserArgAsPPro-GlYProProGlYAlaInglYProAlaIn 418
QY 674 ATGGGGCATCTACGCCCAACCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 733
Db 418 YProGlYglYserArgAsPProGlYProProGlYAlaInglYProAlaInglYProGlYgl 438
QY 734 AGGAACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 778
Db 438 YserArgAsPProGlYProProGlYAlaInglYProAlaInglYProGlYglYserArgAs 458
QY 779 TCTGCGCACACAGCT---CATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 835
Db 458 YProGlYProProGlYAlaInglYProAlaInglYProGlYglY---AlaInglYProAl 477
QY 836 CGCTGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
Db 477 AglYProGlYglYAlaInglYProAlaInglYProGlYglYAla----- 491
QY 896 CCGGCTCCACCATGCTCAACGAGCGAGCGCTGTGCTGCGCGCGCGCGCGCGCG 955
Db 492 -----HisGlYProAlaInglYProGlYglYAla-----Gln-GlYProAlaInglYPro 506
QY 956 GGTGCAAGCGCTTCCCGCAACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
Db 507 GlYglYserArgAsPProGlYProProGlYAlaInglYProAlaInglYProGlYglYser 526
QY 1016 GCAATCAAGTACGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075
Db 527 ArgAsPProGlYProPro-GlYAlaInglYProAlaInglYProAlaInglYProGlYglY 540
QY 1076 TGGAGAACCGGACGCTGAGCGCGCATCTGCGCTGCG-----CCGCGACATGA 1123
Db 540 oGlYglYserArgAsPProGlYProProGlYAlaInglYProAlaInglYProGlYglYse 560
QY 1124 TCGGCTCGAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1160
Db 560 rArgAsPProGlYProProGlYAlaInglYProAlaInglYProGlYglYserArgAsP 580
QY 1161 -TCCGCGCGCTTGGGCGACTTCCCGGGGGGCGTGGGGCGCGCGCGCGCGCGCG 1219
Db 580 oGlYProProGlYAlaInglYProAlaInglYProGlYglY---SerArgAsPProGlYPr 599
QY 1220 TGT- -----CCCGCTGAGAGCGCGCGGTATGAGAGTACCGCGCGCGCGCG 1267
Db 599 oProGlYAlaInglYProAlaInglYProGlYglYserArgAsPProGlYPro----- 616

```

---

```

QY 1268 CGCGCTTCGCGCTGCGCGCGCGCGCGAGCTGACAGCGCGCGGTACTTCCCGCGCGCGTGTG 1327
Db 617 -----ProGlY---AlaInglYProAlaInglY-----ProGlYgl 627
QY 1328 TGTTCGA-----CCGCGCGCA-GGTGGCGCGATACCGCGCGCGCGCGCGCGCTTCCAGACCGCT 1374
Db 627 YserArgAsPProGlYProProGlYAlaInglYProAlaInglYProGlYglYserArgAs 647
QY 1375 ACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1434
Db 647 ProGlYProProGlYAlaInglYPro-----AlaInglYProGlYglYserArg 663
QY 1435 CAGCGCGTTCACCG-----GCCAGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1486
Db 663 gAsPProGlYProProGlYAlaInglYProAlaInglYProGlYglYserArgAsPProGl 683
QY 1487 -GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1545
Db 683 YProProGlYAlaIn-----GlYProAla-GlYProGlYglYserArgAsPProGlYPr 701
QY 1546 ACCGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1602
Db 701 roProGlY-----AlaInglYProAlaInglYProGlYglYAlaInglYProAlaIn 718
QY 1603 GATATCGTGG-----CAAGAGAGTACGCGCGCGCGCGCGCGCGCGCG 738
Db 718 lYProGlYglYAlaInglYProAlaInglYProGlYglYAlaInglYProAlaInglYPro 758
QY 1642 GCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1685
Db 738 YglYAlaInglYProAlaInglYProGlYglYserArgAsPProGlYProProGlYAlaIn 758
QY 1686 -----GCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1725
Db 758 InglYProAlaInglYProGlYglYserArgAsPProGlYProProGlYAlaInglYPro 778
QY 1726 CTGGGCGCA 1733
Db 778 lAglYPro 780

```

---

```

RESULT 7
US-07-757-022B-14
Sequence 14, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114

```

FILING DATE: 29-JUN-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/457,196  
 FILING DATE: 29-DEC-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/390,901  
 FILING DATE: 08-AUG-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Cseri, Luan  
 REGISTRATION NUMBER: 31,822  
 REFERENCE/DOCKET NUMBER: GI 5190  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)876-1170  
 TELEFAX: (617)876-5851  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 941 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-757-022B-14

Alignment Scores:  
 Pred. No.: 6,45E-12 Length: 941  
 Score: 285.50 Matches: 137  
 Percent Similarity: 33.50% Conservative: 60  
 Best Local Similarity: 23.30% Mismatches: 240  
 Query Match: 8.65% Indels: 151  
 Gaps: 24

US-10-009-782-1 (1-1758) x US-07-757-022B-14 (1-941)

32 CCAATGCCAATCCGATTCACAGCCCTTCGACCTGCTCGCGGGGCGACCCCTACG 91  
 167 ProthProthThrThrLysSerAlaProThrThrProLysGluProAlaProThrThr 186  
 92 ACGGACGA-----ACACCCGGGGGGGGCGCCGACCTGGGGCGGCGAC 142  
 187 ThrLysSerAlaProThrThrProLysGluProAlaProThr-----Thr 201  
 143 GCATGCCGCCATCGGCGATCTGCGACGCGCGCGACACCCGGGTGACGTGCG 202  
 202 ThrLysGluProAlaPro-----ThrThrProLysGluProAlaProThr----- 216  
 203 GCGTGTGTCGCGCGCGCTTCGACCTCGACACCCAGCAGCACTACTGCTCA 262  
 217 -----ThrThrLysGluProAlaProThrThrThrLysSer 228  
 263 GGGGTGCGACATGAGCGCCAGATCTCGCAGGGCGCTCACACGCTGCTCGCGCAAT 322  
 229 AlaProThrThr-----ProLysGluProAlaPro----- 238  
 323 GCGGATCAGCTGCGCGCGCTGCGCAGCCACCCGCGCCCTCGACCTGCTG 382  
 239 -----ThrThrProLysGluProAlaProThrThrProLys 250  
 383 ACGAAGCGGCTCTTACGCTTCGAGCGCTTCGCGACTACCTGAGCGGTGCGGCA 442  
 231 GluProAlaProThrThrProLysGluProThrThrThrProLysGluProAlaPro 270  
 443 GCGCGGCGG-----CCCTCAACGCGCGCTGTATGGTGGCCATTCAAGCTGCGCG 493  
 271 ThrThrLysGluProAlaProThrThrPro----- 280  
 494 CCGCGGTATCGCGGACTTGACGCGCGCCAGCAGAGGAATCGCGGCGCATCGGG 553  
 281 -----LysGluProAlaProThrThrAlaProLysLysProAlaPro 293  
 554 ACCTGGCCGAGAGCCATGCGCCAGCGCGCATTCGACGCGCGCTCTTAC 613  
 294 ThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 311  
 614 CGCCGCGCGCGCGCGCA-----CCACCGAAGAGATCATGAGGTGCGCGCGCTGACG 670

312 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrLysSerAla 331  
 671 CGCATGGCGGATCAGCCACCCACATGCGCGAGAGAGGCGACATCTGCGCGCG 730  
 332 ProthThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 351  
 731 TGGAGAACCTCCGATCCGCGCGCGAGCTGAGCGTGGCGGTATCGGACACA 790  
 352 GluProSerProThr-----ThrThrLysGluProAlaProThrThr 365  
 791 AGGTCATGGCGGACCCCATTTGCGCGCGCGCGAGACGCTCCGCTGATCGAGCGC 850  
 366 ProLysGluProAlaPro-----ThrThrPro 374  
 851 CCATGGCGCGGAGAGGTGCTGAGCGGTATCCCTACGTGCGCGCTCCACATGC 910  
 375 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 394  
 911 TCAAGCAGAGACCGGCTGCTGCGCGAGCACCACATCATCCTGTGTGAGAGCGCTTC 970  
 395 ProthThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrThrPro 414  
 971 CCGAATGAGCGGGCGGCGACCTGATGATGCGCGCGCAATCCAGATACG 1030  
 415 LysGluThrAlaProThrThrProLysLysLeuThrProThrThrProLysLysLeuAla 434  
 1031 ACGGTGTCGCGGCTGACGCGCGCGCGCATCTCATCATGATGAGAGAGAGAGAGAG 1090  
 435 ProthThrPro-----GluLysProAlaProThrThrProLysLysLeuAlaProThr 452  
 1091 TCGAGCGCATCTGCGGTGCGCGCGCATGATGCG-----GCTCCGAGCGCTGCGCGCAG 1147  
 453 ThrProLysGluProThrProThrProThrProLysGluProAlaProThrThr----- 469  
 1148 ACGAGCGCGCGCATCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1207  
 469 ----- 469  
 1208 GCGACTGCGCGCTGTCGCGTGAAGAGCGCGGTATGAGATGACCGGCTGACCGCG 1267  
 470 -----Pro 470  
 1268 CCGGCTTCG 1327  
 471 LysAlaAlaAla-----ProSerThrProLysGluProAlaProThrThrProLysGluPro 489  
 1328 TGTTCGACCG 1384  
 490 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 509  
 1385 CCGCGGATCCATTCCGTACGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1444  
 510 ProLysGluThrAlaPro-----ThrThrLysGluProAlaProThrThrProLysLys 528  
 1445 CCGCGGATCCCG 1504  
 529 ProAla-----ProLysGluLeuAlaProThrThrThrLysGluProThrSerThr----- 545  
 1505 TACAAATCGCGGTGAAGCGGGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1564  
 546 -----ThrSerAspLysProAlaProThrThrProLysGluThrAlaPro 560  
 1565 ACATGCGCTCCCTCCCTCGCTGCAATAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624  
 561 ThrThrProLysGluProAlaProThrThrPro----- 571  
 1625 TGGGCGCGCGCTGCG 1684  
 572 -----LysGluProAlaProThrThrProLysGluThrAlaPro-----ThrThrLysLys 588  
 1685 AGGCGAGCG-----GCATGCGGTCTCGACCGCTGTCAGAG-----CCGAGCTGCGCG 1732





```

Yy 1148 ACGAGCGCCGCGCATCCGGCCTGTGGGGACACTTCCCGGGGTGCTGGGCACTATTCGC 1207
Db 550 ----- 550
Yy 1208 GCGACCTGGGCTGTTCCTCCGTGGAGAAGGGGTAATGAATGACGGCTGACC GC 1267
Db 551 -----Pro 551
Yy 1268 CGCGCTTGCGCTGCCGGGCGGGGACGTGACGGCGGCTACTTGGCCGACCTGTGG 1327
Db 552 LysAlaIaIaIaIa-----ProAsnThrProLysIubProAlaProThrThrProLysGIubPro 570
Yy 1328 TGTTCAGCCGGCGGAGGGTGGCCGATACCG----CACCTTGAAACCCCTAACGAGCG 1384
Db 571 AlaProThrThrProLysGIubProAlaProThrThrProLysGIubThrAlaProThrThr 590
Yy 1385 CCGCGCGGATCCATTCCGTGTACGTCAAGCGGGCGGGGTGTGCAAGACGCGTTCA 1444
Db 591 ProLysIubThrAlaPro-----ThrThrLeuLysIubProAlaProThrThrProLysLys 609
Yy 1445 CCGGCCAAGCATGCCGGCGCGGTCTGCAAGCAAGCGCGCGCTGAGCCGGCGGCGGCGCT 1504
Db 610 ProAla-----ProLysGIubLeuAlaProThrThrThrLysGIubProThrSerTh--- 626
Yy 1505 TACAATCCGGCGGAGACGGGGCGCGCTGCGCCGCCCTCCACACCTTGAGCAAACCGCT 1564
Db 627 -----ThrSerAlaPylsProAlaProThrThrThrProLysGIubThrAlaPro 641
Yy 1565 ACATGCCCCCTCCCTCCGCGCGCAATACGGCCACCCGATATCGTGGCAAGAACTGA 1624
Db 642 ThrThrProLysIubProAlaProThrThrPro----- 652
Yy 1625 TGGGCGCGCGCTGCGCGCGGAGCGCAAGGCGGAAAATGACCTGCAAGACTGTGC 1684
Db 653 -----LysGIubProAlaProThrThrThrProLysGIubThrAlaPro-----ThrThrLeuLys 669
Yy 1685 AGGCGACG---GCATCGCGGCTGTGACACCTGTCCAAG-----CGAGCTGGGCG 1732
Db 670 GIubProAlaProThrThrProLysLysIubProAlaProLysGIubLeuAlaProThrThrThr 689
Yy 1733 AGATCGCCCTGAGCTACGAGAAGC 1756
Db 690 LysGIubProThrSerThrThrSer 697

RESULT 9
US-07-757-022B-74
; Sequence 74, Application US/07757022B
; Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
City: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502

```

```

1      FILING DATE: 18-JAN-1991
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: US 07/546,114
4      FILING DATE: 29-JUN-1990
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: US 07/457,196
7      FILING DATE: 29-DEC-1989
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: US 07/390,901
10     FILING DATE: 08-AUG-1989
11     ATTORNEY/AGENT INFORMATION:
12     NAME: Cseri, Luann
13     REGISTRATION NUMBER: 31,822
14     REFERENCE/DOCKET NUMBER: GI 5190
15     TELECOMMUNICATION INFORMATION:
16     TELEPHONE: (617)876-1170
17     TELEFAX: (617)876-5851
18     INFORMATION FOR SEQ ID NO: 74:
19     SEQUENCE CHARACTERISTICS:
20     LENGTH: 1038 amino acids
21     TYPE: AMINO ACID
22     TOPOLOGY: linear
23     MOLECULE TYPE: protein
24     US-07-757-022B-74

```

Alignment Scores:	
Pred. No.:	6,61e-12
Score:	285.50
Percent Similarity:	33.50%
Best Local Similarity:	23.30%
Query Match:	8.05%
DB:	
	4
	Length: 1038
	Matches: 137
	Conservative: 60
	Mismatches: 240
	Indels: 151
	Gaps: 24

```

QY 32 CAGTGTCCCAATCCGATTCCCGACCCCTTGACCTGCTGCTCGGGGGGCAACCTCATCG 91
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 ProthrProthrThrThryLysSerAlaProthrThrProlyGluProAlaProthrThr 25
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 92 ACGGCAGCA-----ACACCCGGGGGGGGGGCGGCGACCTGGGCGTGGCGGCGACC 144
   ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 ThrLysSerAlaProthrThrProlyGluProAlaProthr-----Thr 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 GCATGCCCGCATCGCGCATGTCTGCGACGGCGCGCGCACACCGGGGTGACGTGCGG 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 267 ThrLysGluProAlaPro-----ThrThrProlyGluProAlaProthr----- 281
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 203 GCCTGTGTGTGCGCGCGGCTTCATCGACTGCGACACCGACGACAGCAACTGCTGTCA 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 282 -----ThrThryGluProAlaProthrThrThryLysSer 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 263 GGGCTGCGGACATGACGCCCAAGATCTGCGAGGGCGCTCACCGAGTGTCACGGGCAATT 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 AlaProthrThr-----ProlyGluProAlaPro----- 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 323 GCGGCATCAGCCTGGCGCGGCTGGGGGACCGCGCAACCGCGCCCGCCCTGAGACTGTGG 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 -----ThrThrProlySerProAlaProthrThrProlys 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 383 ACGAAGCGGCTCTTACGCTTTCAGAGCGCTTCGGCGCATCTGAGACCGGTTGCGGGCCA 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 GluProAlaProthrThrProlyGluProthProthrThrProlyGluProAlaPro 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 443 CGCGGGCGG-----CGGTCAACGCGCGCTGTATGTGGGCGCATTCACGCTGCGG 493
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 ThrThryGluProAlaProthrThrPro----- 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 494 CCGGGGTGATCCCGGAATTGACAGCGCGCGCCACCGACGAGAAATCGGCGCATGCGGG 553
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 -----LysGluProAlaProthrAlaProthrProlySerProAlaPro 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 554 ACCTGCGGAGGAACCATGGCCACAGCGCGCCCATCGGCAATTTCGACCGGCGCTTCTACC 613
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 359 ThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 376
QY 614 GCGCCGCGCGCGCGCGCA---CCACGAGAGATCATGAGTGTGCGCGCGCTGAGCG 670
Db 377 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrThrLysSerAla 396
QY 671 CGCATGGCGGCGCATGACGCGCACCATGCGCGGAGAGCGGACATGTCGCGCGCG 730
Db 397 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 416
QY 731 TGGAGAAACCTTCGCGCATGCGCGGAGCGGAGCGGCGGATGATCTCCACACCA 790
Db 417 GluProSerProThr-----ThrLysGluProAlaProThrThr 430
QY 791 AGGTCATGGGCGACGCAATTTGCGCGCTGCGCGGAGCGCTGCGCTGATGAGGCGG 850
Db 431 ProLysGluProAlaPro-----ThrThrPro 439
QY 851 CCATGGGCGCGCGAGAGCTGCTGCGTGGAGCGGATACCTACGTCGCGCGCTCCACCATG 910
Db 440 LysLysProAlaProThrThrThrProLysGluProAlaProThrThrProLysGluProAla 459
QY 911 TCAGAGAGAGCGCGCTGCTGCGCGGAGCGACCATCATCACTGTCGTCGACGCGCTTCC 970
Db 460 ProThrThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrThrPro 479
QY 971 CCGAAGTGAAGCGGCGCGCGAGCGGATGAGTGGCGCGCGCGCGCAATTCAGATAGC 1030
Db 480 LysGluThrAlaProThrThrProLysLysLeuThrProThrProGluLysLeuAla 499
QY 1031 AGGTGTCGCGCGAGCTGAGCGCGCGCGCGCTACTACTCATGATGAGACCAACCGAGC 1090
Db 500 ProThrThrPro-----GluLysProAlaProThrThrProGluLysLeuAlaPro 517
QY 1091 TGCACGCGATCTGCGCTGCGCGCGCGCGAGCATGATG---GCTCGGAGCGCTGCGCGCAG 1147
Db 518 ThrProGluLysProThrProThrThrProGluLysProAlaProThrThr----- 534
QY 1148 ACGAGCGCGCGCGATCCGCGCTGCGGCGCGCTTCCGCGGCTGCGGCGCATATGCGC 1207
Db 534 ----- 534
QY 1208 GCGACCTGCGGCTGTCCCGCTGAGAGCGCGGATGAGATGACCGCGCTGACGCGC 1267
Db 535 ----- 535
QY 1268 GCGCCTTCGCGCTGCGCGCGCGCGGAGCGTGCAGCGCGGCTACTTGCAGCGCTGTG 1327
Db 536 LysAlaAlaAla---ProAlaThrProLysGluProAlaProThrThrProLysGluPro 554
QY 1328 TGTTCGACCGCGCGAGCGTGCATACG---CCACCTTCGAAACCCCTACCGAGCGCG 1384
Db 555 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 574
QY 1385 CCGCGCGCATCATTCGCTAGCTCAAGCGCGCGCGCTGCGCAAGAGAGCGCTGCA 1444
Db 575 ProLysGluThrAlaPro---ThrThrLysGluProAlaProThrThrProLysLys 593
QY 1445 CCGGCGAGCATGCGCGCGCGCTGTCGCAAGCGCGCGCGCTGAGCGCGCGCGAGCCCT 1504
Db 594 ProAla-----ProLysGluLeuAlaProThrThrLysGluProThrSerThr----- 610
QY 1505 TACAATCCGCGGTAAGAGGCGCGCGCTGCGCGCTCCCGCAACCTGGAGCGAAACCGCT 1564
Db 611 ----- 611
QY 1565 ACATGCGCGCGCTTCGCTGCGCATACGAGCGCGCGCGCATATGCTGGGCAAGAGATGA 1624
Db 626 ThrThrLysGluProAlaProThrThrPro----- 636
QY 1625 TGGGCGCGCGCTGCGCGCGCGAGCGCAAGCGCGGAAATGACCTGCAACACCTGTGCG 1684
Db 637 -----LysGluProAlaProThrThrProLysGluThrAlaPro---ThrThrLysLys 653

```

```

QY 1685 AGCGCAGCG---GCATGCGGCGTCTGACCGCTGTCCAAG-----CCGAGCTGGGCC 1732
Db 654 GluProAlaProThrThrProLysLysProAlaProLysGluLeuAlaProThrThr 673
QY 1733 AGATCCCGCTGAGCTACGAGAGC 1756
Db 674 LysGluProThrSerThrThrSer 681

RESULT 10
US-07-757-022B-58
; Sequence 58, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseiff, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-58

Alignment Scores:
Pred. No.: 6.63e-12 Length: 1049
Score: 285.50 Matches: 137
Percent Similarity: 33.50 Conservative: 60
Best Local Similarity: 23.30 Mismatches: 240
Query Match: 8.65e Indels: 151
DB: 4 Gaps: 24

US-10-009-782-1.(1-1758) x US-07-757-022B-58 (1-1049)

```



Alignment Scores:	
Pred. No.:	6,77e-12
Score:	285.50
Percent Similarity:	33.50%
Best Local Similarity:	23.30%
Query Match:	8.65%
DB:	4
	Length: 1140
	Matches: 157
	Conservative: 60
	Mismatches: 240
	Indels: 151
	Gaps: 24

[illegible]

QY 1621 TGGGGCGCGCTGGCGCGAGCGAGCGCGGAAATGACCTGCAAGACCTGTGCC 1684  
 Db 771 -----LysGluProAlaProThrThrProLysGlyThrAlaPro---ThrThrLeuLys 787  
 QY 1685 AGCCAGCG--GCATCGCGGTCTCGACCTGTCCAGG-----CCGAGCTGGGCG 1732  
 Db 788 GluProAlaProThrThrProLysProAlaProLysGluLeuAlaProThrThrThr 807  
 QY 1733 AGATCGCCCTGATGATCGAGAAC 1756  
 Db 808 LysGlyProThrSerThrThrSer 815  
 RESULT 12  
 US-07-757-022B-44  
 Sequence 44, Application US/0757022B  
 Patent No. 6433142  
 GENERAL INFORMATION:  
 APPLICANT: Gesner, Thomas G.  
 APPLICANT: Clark, Stephen C.  
 APPLICANT: Turner, Katherine  
 APPLICANT: Hewick, Rodney M.  
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 NUMBER OF SEQUENCES: 143  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genetics Institute, Inc.  
 STREET: 87 Cambridge Park Drive  
 CITY: Cambridge  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02140  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/757,022B  
 FILING DATE: 19910910  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/643,502  
 FILING DATE: 18-JAN-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/546,114  
 FILING DATE: 29-JUN-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/457,196  
 FILING DATE: 29-DEC-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/390,901  
 FILING DATE: 08-AUG-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Geert, Luann  
 REGISTRATION NUMBER: 31,822  
 REFERENCE/DOCKET NUMBER: GI 5190  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)876-1170  
 TELEFAX: (617)876-5851  
 INFORMATION FOR SEQ ID NO: 44:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1270 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-757-022B-44  
 Alignment Scores:  
 Pred. No.: 6.95e-12 Length: 1270  
 Score: 285.50 Matches: 137  
 Percent Similarity: 33.50% Conservative: 60  
 Best Local Similarity: 23.30% Mismatches: 240  
 Query Match: 8.65% Indels: 151  
 Db: 4 Gaps: 24

US-10-009-782-1 (1-1758) x US-07-757-022B-44 (1-1270)  
 QY 32 CCATGTCCTCCATTCGATCCAGCCCTTCGACCTGTCTGCGGGGCGACCTCATGC 91  
 Db 232 ProThrProThrThrThrLeuLysSerAlaProThrThrProLysGluProAlaProThrThr 251  
 QY 92 AGCGGAGCA-----ACACCCCGGGGGCGCGCGCGGACCTGGCGGGCGAGC 142  
 Db 252 ThrLysSerAlaProThrThrProLysGluProAlaProThr-----Thr 266  
 QY 143 GCATCGCCGATCGCGGATCTGTGAGACCGCGCGCGCACACCGGGGTGAGGTGCGG 202  
 Db 267 ThrLysGluProAlaPro-----ThrThrProLysGluProAlaProThr----- 281  
 QY 203 GCGTGTGTGTGCGCGCGCGGCTTCATCGACTGTGACACCCAGCAGCAGCACTACTGTCTCA 262  
 Db 282 -----ThThrLysGluProAlaProThrThrThrLysSer 293  
 QY 263 GCGTCGCGCATGATGACGCCCAAGATCTCGCAGGGCGGTACACGAGGTGTACAGGCGCAT 322  
 Db 294 AlaProThrThr-----ProLysGluProAlaPro----- 303  
 QY 323 GCGGATCAGCTGCGCGCGCGCTGCGCGCACACCGCGCGCGCGCTGAGCTGTG 382  
 Db 304 -----ThrThrProLysProAlaProThrThrProLys 315  
 QY 383 ACGAAGCGCGCTTACCGTTTGAGGCGCTGCGCGCATCTGACCGCGTGGCGGCA 442  
 Db 316 GluProAlaProThrThrProLysGluProThrProThrThrProLysGluProAlaPro 335  
 QY 443 CGCGGCGG-----CGGTACAGCGCGCGGTATGTGTGCGCATTAAGCTGCGG 493  
 Db 336 ThrThrLysGluProAlaProThrThrPro----- 345  
 QY 494 CGCGGTCATGCGGACTGTGACGCGCGCGCGCACCGACGAGAAATCGCGCATGCGG 553  
 Db 346 -----LysGluProAlaProThrAlaProLysProAlaPro 358  
 QY 554 ACCTGCGCGAGGAGGACCATGCGCGCGCGCATGCGGCTTCGACCGCGGCTCTAC 613  
 Db 359 ThrThrProLysGluProAlaProThrThrProLys-----GluProAlaProThrThr 376  
 QY 614 CGCGCGCGCGCGCGCA---CCACGGAAGATCATGAGGTGTGCGCGCGCTGACG 670  
 Db 377 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrThrLysSerAla 396  
 QY 671 CGCATGCGCGCATACGCGCACCAATGCGGAGAGAGCGAGCATCTGTGCGCGG 730  
 Db 397 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 416  
 QY 731 TGGAGAACTTCGCGCATGCGCGCGCGAGCTGACGTCGCGGTGATCTGCGACCA 790  
 Db 417 GluProSerProThr-----ThrThrLysGluProAlaProThrThr 430  
 QY 791 AGTATGAGCGGCAATTCGCGCGCGCGCGGAGACCGCTGCGGTGATCGAGCGG 850  
 Db 431 ProLysGluProAlaPro-----ThrThrPro 439  
 QY 851 CCATGCGCGCGCGCGCTGCTGCGAGCGGTATCCCTGAGCGCGGCTCCACCATGC 910  
 Db 440 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 459  
 QY 911 TCAAGCAGAGCGCGGTGCTGCGCGCGAGCAGCATCATCATCTGTGCAAGCCCTTC 970  
 Db 460 ProThrThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrThrPro 479  
 QY 971 CCGAATGAGCGCGCGCGACCTGATGATGAGCGCGCGCGGCGGCAATTCAGTACG 1030  
 Db 480 LysGluThrAlaProThrThrProLysLysLeuThrProThrProLysLysAla 499  
 QY 1031 ACGTGTGCCCGAGCTCAGCGCGCGCGCATCTACTATCATGATGAGAACCGGAGC 1090  
 Db 1031 -----



Db 500 ProThrPro-----gluysProAlaProThrThrProgluLeuAlaProThr 517  
 QY 1091 TCGAGCCGATCGGCGCTGGCGCCGACATGATCG---GCTCCGACGCGCTCCGACG 1147  
 Db 518 ThrProgluLupProThrProThrProThrProgluLupProAlaProThr----- 534  
 QY 1148 ACAGAGCCCGGATCGCGCTGGGGACCTTCCCGGGGTGTGGGGCACTATGCGC 1207  
 Db 534 ----- 534  
 QY 1208 GCGACCTGGGCTGTCCGCTGGAGACGGGTATGAAATGACGGCGCTGACGCGCG 1267  
 Db 535 -----Pro 535  
 QY 1268 CCGGCTTGGCGCTGGCGCGGCGGACCTGACGCGCGGCTGCTGCGGACCTGGTGG 1327  
 Db 536 LysAlaAlaAla---ProsnThrProLysgluProAlaProThrThrProLysgluPro 554  
 QY 1328 TGTGACCGCGGACGAGGCGCGATACCG---CCACCTTCGACACCTTACCGAGCGG 1384  
 Db 555 AlaProThrThrProLysgluProAlaProThrThrProLysgluThrAlaProThrThr 574  
 QY 1385 CCGCGCGCATCCATCCGTCGTACGTCACAGCGCGCGCTGCTGCGACAGAGCGGCTCA 1444  
 Db 575 ProLysgluThrAlaPro---ThrThrLeuLysgluProAlaProThrThrProLysLys 593  
 QY 1445 CCGGCGACGATCGCGCGCGCTGCTGCGACGACGCGCGCGCTGAGCGCGCGCGCGCT 1504  
 Db 594 ProAla---ProLysgluLeuAlaProThrThrThrLysgluProThrThr--- 610  
 QY 1505 TACATCCGCGGTGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1564  
 Db 611 -----ThrsrAspLysProAlaProThrThrProLysgluThrAlaPro 625  
 QY 1565 ACATGCGCCCTCCCTCGCTGCGATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624  
 Db 636 ThrThrProLysgluProAlaProThrThrPro----- 636  
 QY 1625 TGGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1684  
 Db 637 -----LysgluProAlaProThrThrProLysgluThrAlaPro---ThrThrLeuLys 653  
 QY 1685 AGCGCAGCG---GCATCGCGGCTGTCGACCTGTCGACG---CCGACCTGGCGC 1732  
 Db 654 GluProAlaProThrThrProLysLysProAlaProLysgluLeuAlaProThrThr 673  
 QY 1733 AGATCGCGCGAGCTAGAGAGAC 1756  
 Db 674 LysgluProThrThrThrSer 681

RESULT 13  
 US-07-757-022B-42  
 Sequence 42, Application US/07757022B  
 Patent No. 6433142  
 GENERAL INFORMATION:  
 APPLICANT: Geesner, Thomas G.  
 APPLICANT: Clark, Stephen C.  
 APPLICANT: Turner, Katherine  
 APPLICANT: Hewick, Rodney M.  
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 NUMBER OF SEQUENCES: 143  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genetics Institute, Inc.  
 STREET: 87 Cambridgepark Drive  
 CITY: Cambridge  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02140-

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/757,022B  
 FILING DATE: 19910910  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/643,502  
 FILING DATE: 18-JAN-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/546,114  
 FILING DATE: 29-JUN-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/457,196  
 FILING DATE: 29-DEC-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/390,901  
 FILING DATE: 08-AUG-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gserr, Luann  
 REGISTRATION NUMBER: 31,822  
 REFERENCE/DOCKET NUMBER: GI 5190  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)876-1170  
 TELEFAX: (617)876-5851  
 INFORMATION FOR SEQ ID NO: 42:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1311 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-757-022B-42

Alignment Scores:  
 Pred. No.: 7.01e-12  
 Score: 285.50  
 Percent Similarity: 33.508  
 Best Local Similarity: 23.308  
 Query Match: 8.658  
 DB: 4  
 Gaps: 24

US-10-009-782-1 (1-1758) x US-07-757-022B-42 (1-1311)

QY 32 CCATGTCACCAATCCGATCCGACCGCTGACCTGCTGCGGCGGACCTCATG 91  
 Db 273 ProThrProThrThrThrLysSerAlaProThrThrProLysgluProAlaProThr 292  
 QY 92 ACAGCAGCA-----ACACCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 142  
 Db 293 ThrLysSerAlaProThrThrProLysgluProAlaProThr-----Thr 307  
 QY 143 GCATGCGCGCATGCGCATGTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 202  
 Db 308 ThrLysgluProAlaPro-----ThrThrProLysgluProAlaProThr----- 322  
 QY 203 GCGTGGTGGTGGCGCGCGCTTCATGCACTGCGACACCGACCGAGAGACATGCTGCA 262  
 Db 323 -----ThrThrLysgluProAlaProThrThrLysSer 334  
 QY 263 GCGCTCCGACATACGCGCCCAAGATCTCGAGAGCGGTACACAGCGGTGTACGCGCAAT 332  
 Db 335 AlaProThrThr-----ProLysgluProAlaPro----- 344  
 QY 323 GCGGATCAGACCTGCGCGCGCTGCGCGACCGCAACCGCGCGCGCGCGCGCGCTG 382  
 Db 345 -----ThrThrProLysLysProAlaProThrThrProLys 356  
 QY 383 ACAGAGCGGCTTACCGTTGAGAGGCTCCGCGACATGACGAGCGGTTGCGGCGCA 442  
 Db 357 GluProAlaProThrThrProLysgluProThrThrProLysgluProAlaPro 376  
 QY 443 CCGCGCGG-----CCGTACAGCGCGCGCTGATGTGGGCAATTCAGCGTGGCGG 493  
 Db 377 ThrThrLysgluProAlaProThrThrPro----- 386

QY 494 CCGCGGTCAATGCGGACTTGCAGCGCGCCGACGCGGAAATCCGCGCCATGCGG 553  
 Db 387 -----LysGluProAlaProThrAlaProLysProAlaPro 399  
 QY 554 ACCTGGCCGAGAGCCATGCGCGCGCCATCGCATTCGACCGGCGCTTCTAC 613  
 Db 400 ThrThrProLysGluProAlaProThrProLys-----GluProAlaProThrThr 417  
 QY 614 CCGCCGCGCGCGCGCA---CCAGGAGAGATCATCGAGTGTCCGCGCGCTAGCG 670  
 Db 418 ThrLysGluProSerProThrThrProLysGluProAlaProThrThrLysSerAla 437  
 QY 671 CGCATGGCGGCACTCAAGCCACCATGCGGAGCGGAGCGGACATCGGCGCGCC 730  
 Db 438 ProThrThrThrLysGluProAlaProThrThrThrLysSerAlaProThrThrProLys 457  
 QY 731 TGGAGGAAACCTTCGCGCATGCGCGCGCGAGCTGCGGTGTGATCTGCACACA 790  
 Db 458 GluProSerProThr-----ThrThrLysGluProAlaProThrThrThr 471  
 QY 791 AGGTCAATGCGGCGCAATTCGCGCGCGTCCGCGAGAGCGTCCGCGCATGACGCCG 850  
 Db 472 ProLysGluProAlaPro-----ThrThrPro 480  
 QY 851 CCATGGCGCGCGAGGAGCTCTGCGTGCATCCATCGTGGCGCGCGCTCCACCATGC 910  
 Db 481 LysLysProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAla 500  
 QY 911 TCAGCAGGACCGCGCTGTGTGTGCGCGGACGACCATCATCATCGTGGACGCCCTTCC 970  
 Db 501 ProThrThrThrLysLysProAlaProThrAlaProLysGluProAlaProThrPro 520  
 QY 971 CCGAATGAGCGGCGGCGACCTGTGATGAGTGGCGGCGGCGGCAATCCAAATGACG 1030  
 Db 521 LysGluThrAlaProThrThrProLysLysLysLysLysLysLysLysLysLysLys 540  
 QY 1031 ACGTGGTGGCGGAGCTGACGCGCGCGCGCGCATCTCATGATGAGCAGCAACCGGAGC 1090  
 Db 541 ProThrThrPro-----GluLysProAlaProThrThrProGluLysLysLysLysLys 558  
 QY 1091 TGCAGCGCATCTGCGGTGCGCGCGGACGATGATG---GCTCGGAGCGCGCTGCGCGCAG 1147  
 Db 559 ThrProGluGluProThrProThrProThrProGluLysLysLysLysLysLysLys 575  
 QY 1148 ACGAGCGCGCGCATCGCGCGCTGTGGGACACTTCCGCGGCTGTGGGACATGATGCGC 1207  
 Db 575 -----Pro 576  
 QY 1208 GCGACTGGGCGCTGTCCGCTGAGAGCGGCGGATGAGAGATGACCGCGCTGACCGCG 1267  
 Db 576 -----Pro 576  
 QY 1268 CCGGCTTGGCGCTGGCGCGCGCGGAGCTGAGCGCGGCTACTTGGCGGAGCTGTGG 1327  
 Db 577 LysAlaAlaAla---ProAlaThrProLysGluProAlaProThrThrProLysGluPro 595  
 QY 1328 TGTTCAGACCGCGGACGAGTGGCGGATGCG---CCACTTGAACAACCTTACGAGCGG 1384  
 Db 596 AlaProThrThrProLysGluProAlaProThrThrProLysGluThrAlaProThrThr 615  
 QY 1385 CCGCGCGCATTCATTCGCGTACGATGAGCGCGCGCGCTGTGCAAGACGAGCGCTTCA 1444  
 Db 616 ProLysGluThrAlaPro---ThrThrLysGluProAlaProThrThrProLysLys 634  
 QY 1445 CCGCGCAGATGCGCGCGCGCTGCTGCACGACGAGCGCGCGCGCGCGCGCGCGCGCCCT 1504  
 Db 635 ProAla-----ProLysGluLysLysLysLysLysLysLysLysLysLysLysLys 651  
 QY 1505 TACAAATCCGCGGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1564  
 Db 652 -----ThrSerLysProAlaProThrThrProLysGluThrAlaPro 666  
 QY 1565 ACATGGCGCGCTCCCTCCGCTCGCAATAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624

Db 667 ThrThrProLysGluProAlaProThrThrPro----- 677  
 QY 1625 TGGCGCGCGCGCTGGCGCGCGCGAGCGCGGAAATGACCTGAGACCTGTGCG 1684  
 Db 678 -----LysGluProAlaProThrThrProLysGluThrAlaPro---ThrThrLys 694  
 QY 1685 AGCGCGCG---GCATCGCGGTCTGACCTCTGTCCAGG-----CCGAGCTGGCGC 1732  
 Db 695 GluProAlaProThrThrProLysLysProAlaProLysGluLysLysLysLysLys 714  
 QY 1733 AGATCGCGCTGAGCAGAGAGC 1756  
 Db 715 LysGluProThrSerThrThrSer 722  
 RESULT 14  
 US-07-757-022B-142  
 ; Sequence 142, Application US/07757022B  
 ; Patent No. 6433142  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gesner, Thomas G.  
 ; APPLICANT: Clark, Stephen C.  
 ; APPLICANT: Turner, Katherine  
 ; APPLICANT: Hewick, Rodney M.  
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 ; NUMBER OF SEQUENCES: 143  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Genetics Institute, Inc.  
 ; STREET: 87 Cambridgepark Drive  
 ; CITY: Cambridge  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02140  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/757,022B  
 ; FILING DATE: 19910910  
 ; CLASSIFICATION: 530  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/643,502  
 ; FILING DATE: 18-JAN-1991  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/546,114  
 ; FILING DATE: 29-JUN-1990  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/457,196  
 ; FILING DATE: 29-DEC-1989  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/390,901  
 ; FILING DATE: 08-AUG-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Casert, Luann  
 ; REGISTRATION NUMBER: 31,822  
 ; REFERENCE/DOCKET NUMBER: GI 5190  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617)876-1170  
 ; TELEFAX: (617)876-5851  
 ; INFORMATION FOR SEQ ID NO: 142:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1313 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-07-757-022B-142  
 Alignment Scores:  
 Score: 7.01e-12 Length: 1313  
 Pred. No.: 285.50 Matches: 137  
 Percent Similarity: 33.50% Conservative: 60

523 LysGluThrAlaProThrThrProIvalIvalLeuThb-D-Asn-L-Val  
||| ||| |||  
|||||

COUNTRY: U.S.A.  
ZIP: 02140

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk



```

Db 655 -----ThSerAspLysProAlaProThrThrProLysGlyThrAlaPro 669
QY 1565 ACATGGCCCCCTCCCTCCGCTCGCAATACGGCCCAACCGATATCGTGGCAAGAGTGA 1624
    ||| ||| |||
Db 670 ThrThrProLysGluProAlaProThrThrPro-----680
QY 1625 TGGGGCGCGGCTGGCGCGGCGCAAGCGCCGGAATAAGACCTGCAAGACCTGTCCG 1684
    ::::: ||| ||| |||
Db 681 -----LysGluProAlaProThrThrProLysGlyThrAlaPro--ThrThrLeuLys 697
    ::::: ||| ||| |||
QY 1685 AGGCCAGCG--GCATCGCGGTCTGACCTGTCTCAAG-----CCGAGCTGGGCG 1732
    ||||| ::::: ||| ||| |||
Db 698 GluProAlaProThrThrProLysProAlaProLysGluLeuAlaProThrThrThr 717
    ::::: ||| ||| |||
QY 1733 AGATCGCCCTGAGCTACGAGAAGC 1756
    ::::: ||| |||
Db 718 LysGlyProThrSerThrThrSer 725

```

Search completed: May 11, 2003, 12:17:20  
 Job time : 65.5 secs

Run on: May 11, 2003, 10:52:45 ; Search time 66 Seconds

May 11, 2003, 10:52:45 ; Search time 66 Seconds  
(without alignments)  
7098.624 Million cell updates/sec

```

Title:      US-10-009-782-1
Perfect score: 3299
Sequence:   1 gaattccacttgatcgcgga.....ccctgagctacgagaagcct 1750

```

Scoring table:			
BLOSUM62			
Xgapop	10.0	Xgapext	0.5
Ygapop	10.0	Ygapext	0.5
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Total number of hits satisfying chosen parameters: 1816940

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:  
-MODEL = freemot -30 -33 -1

[illegible]

Database : A\_Geneseq\_101002:

1:	/SID2/gcgdata/genseq/genseqp_emb1/AA1980.DAT.*
2:	/SID2/gcgdata/genseq/genseqp_emb1/AA1981.DAT.*
3:	/SID2/gcgdata/genseq/genseqp_emb1/AA1982.DAT.*
4:	/SID2/gcgdata/genseq/genseqp_emb1/AA1983.DAT.*
5:	/SID2/gcgdata/genseq/genseqp_emb1/AA1984.DAT.*
6:	/SID2/gcgdata/genseq/genseqp_emb1/AA1985.DAT.*
7:	/SID2/gcgdata/genseq/genseqp_emb1/AA1986.DAT.*
8:	/SID2/gcgdata/genseq/genseqp_emb1/AA1987.DAT.*
9:	/SID2/gcgdata/genseq/genseqp_emb1/AA1988.DAT.*
10:	/SID2/gcgdata/genseq/genseqp_emb1/AA1989.DAT.*
11:	/SID2/gcgdata/genseq/genseqp_emb1/AA1990.DAT.*
12:	/SID2/gcgdata/genseq/genseqp_emb1/AA1991.DAT.*
13:	/SID2/gcgdata/genseq/genseqp_emb1/AA1992.DAT.*
14:	/SID2/gcgdata/genseq/genseqp_emb1/AA1993.DAT.*
15:	/SID2/gcgdata/genseq/genseqp_emb1/AA1994.DAT.*
16:	/SID2/gcgdata/genseq/genseqp_emb1/AA1995.DAT.*
17:	/SID2/gcgdata/genseq/genseqp_emb1/AA1996.DAT.*
18:	/SID2/gcgdata/genseq/genseqp_emb1/AA1997.DAT.*
19:	/SID2/gcgdata/genseq/genseqp_emb1/AA1998.DAT.*
20:	/SID2/gcgdata/genseq/genseqp_emb1/AA1999.DAT.*
21:	/SID2/gcgdata/genseq/genseqp_emb1/AA2000.DAT.*
22:	/SID2/gcgdata/genseq/genseqp_emb1/AA2001.DAT.*
23:	/SID2/gcgdata/genseq/genseqp_emb1/AA2002.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	2511	76.1	484	22	AA848975	Alcaligenes xylosoxidans
2	2481	75.2	484	21	AAAY84943	Amino acid sequence
3	942	28.6	536	22	AAAB96073	Putative P. abyssi
4	612	18.6	558	22	AAAB47456	D-ethanolamine dehydrogenase
5	411	12.5	1212	20	AAH87503	Human N-methyl-D-arginine
6	402.5	12.5	1061	20	AAH87504	Human N-methyl-D-arginine
7	378.5	11.5	1332	21	AAAB26239	Human N-methyl-D-arginine
8	378.5	11.5	1232	21	AAAB47961	Human N-methyl-D-arginine
9	370	11.2	1081	21	AAAB26240	Human NMDA receptor
10	370	11.2	1081	23	ABBS55523	Human N-methyl-D-arginine
11	341	10.5	900	22	ABG03553	Human NMDA receptor
12	339	10.3	1592	22	ABBS5827	Human NMDA receptor
13	332.5	10.1	999	22	ABBS5817	Human NMDA receptor
14	331.5	10.0	1515	22	ABBS5826	Human NMDA receptor
15	321.5	9.7	599	22	ABG03530	Human NMDA receptor
16	318	9.8	506	22	AAAY04841	Human NMDA receptor
17	318	9.8	1592	22	ABBS5827	Human NMDA receptor
18	313.5	9.7	599	22	ABG03530	Human NMDA receptor
19	312	9.5	558	20	AAI22184	Human NMDA receptor
20	309	9.4	900	22	ABG03533	Human NMDA receptor
21	307.5	9.4	999	22	ABBS5817	Human NMDA receptor
22	307	9.3	1017	22	AAAB59813	Human NMDA receptor
23	306.5	9.3	572	22	AAAY04954	Human NMDA receptor
24	302	9.2	627	22	AAU05180	Human NMDA receptor
25	295	9.1	1615	22	ABBS5826	Human NMDA receptor
26	294	9.1	1061	20	AAH87504	Human NMDA receptor
27	290.5	8.8	1098	22	ABG20365	Human NMDA receptor
28	289.5	8.8	5179	22	AAAB24515	Human NMDA receptor
29	287.5	8.9	1212	20	AAH87503	Human NMDA receptor
30	286	8.8	1081	21	AAAB26240	Human NMDA receptor
31	286	8.8	1081	21	ABBS55523	Human NMDA receptor
32	285.5	8.8	1017	22	ABBS59813	Human NMDA receptor
33	285.5	8.7	1299	22	AAH47332	Human NMDA receptor
34	285.5	8.7	1404	23	AAH26049	Human NMDA receptor
35	285.5	8.7	1404	23	AAH26058	Human NMDA receptor
36	285.5	8.7	1404	22	AAAB29773	Human NMDA receptor
37	285.5	8.7	1415	22	AAU32262	Human NMDA receptor
38	284.5	8.6	476	22	ABG03460	Human NMDA receptor
39	283	8.6	440	22	ABG05166	Human NMDA receptor
40	283	8.6	819	22	ABG03559	Human NMDA receptor
41	283	8.6	1065	14	AAH37745	Human NMDA receptor
42	283	8.6	1065	14	AAH37745	Human NMDA receptor
43	283	8.6	1065	19	AAH37745	Human NMDA receptor
44	277	8.4	1605	22	AAAB59824	Human NMDA receptor
45	276	8.4	1040	22	ABG14734	Human NMDA receptor

## ALIGNMENTS

1

	AAB489/5 standard; protein; 484 AA
ID	
XX	

AC AAB48975

27-MAR-2001 (first entry)

XX 22

KW D-amino acid  
KW 21no-oxh

KW: antibiotics

AA  
OS

XX

XX 13

PD 28-DEC-20



XX 15-JUN-2000; 2000MO-JP03932.  
 XX 17-JUN-1999; 99JP-0170555.  
 XX (AMANO) AMANO ENZYME INC.  
 XX Takeuchi K, Koide Y, Hirose Y, Moriyuchi M, Isobe K;  
 DR WPI; 2001-080828/09.  
 DR N-PSDB; AAC91797.

XX Transformed microorganism from zinc-tolerant host for selective  
 PT production of D-aminoacylase, useful in synthesis of high  
 PT optical-purity D-amino-acids for antibiotic side-chains and peptide  
 PT drugs -

XX Claim 2; Page 14-18; 22pp; Japanese.

CC The invention relates to a recombinant zinc-tolerant microorganism which  
 CC expresses the D-aminoacylase from Alcaligenes xylosoxidans subspecies  
 CC xylosoxidans. The presence of zinc ions in the culture medium enhances  
 CC expression of the D-aminoacylase gene, and the invention also relates to  
 CC the process of recombinantly producing the D-aminoacylase using the  
 CC microorganism of the invention. The recombinant microorganism is used  
 CC for the selective production of D-aminoacylase, which is useful in the  
 CC synthesis of high optical-purity D-form amino acids for use in the  
 CC production of antibiotics and peptide drugs. The present sequence  
 CC represents Alcaligenes xylosoxidans subspecies xylosoxidans  
 CC D-aminoacylase.

XX Sequence 484 AA;

Alignment Scores:

Pred. No.: 4.43e-173 Length: 484  
 Score: 2511.00 Matches: 484  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 76.11% Indels: 0  
 DB: 22 Gaps: 0

US-10-009-782-1 (1-1758) x AAB48975 (1-484)

QY 34 ATGTCCCAATCCGATTCAGACCCCTTGCAGCTGCTGCTCCGGGCGGCACCCCTCATGAC 93  
 DB 1 MetSerGlnSerAspSerGlnProPheAspLeuLeuAlaGlyThrLeuIleAsp 20  
 QY 94 GGCAGCAACACCCCGGGGCGCGCGCGACCTGGCGTGGCGCGCGCGACCGCATCGCGCC 153  
 DB 21 GlySerAsnThrProGlyArgArgAlaAspLeuGlyValArgGlyAspArgIleAlaAla 40  
 QY 154 ATGGCGATCTGTGCGAGCGCGCGCGCGCGACACCCGGGTGCAGCTGTGGGCGCTGTGTC 213  
 DB 41 IleGlyAspLeuSerSerPalaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaVal 60  
 QY 214 GCGCCCGGCTTCATCGACCTCGACACCGACGAGACACTACTGCTGAGGGGTGCGCGAC 273  
 DB 61 AlaProGlyPheIleAspSerHisThrHisAspAspAspTrpLeuLeuArgArgAsp 80  
 QY 274 ATAGGCGCAAGATCTCGAGAGCGGTGCACACCGGTGTGCAGGGAATTCGCGCATCAGC 333  
 DB 81 MetThrProLysIleSerIleGlyValThrThrValIleThrGlyAsnGlySerIleSer 100  
 QY 334 CTGGCGCGCTGGCG 393  
 DB 101 LeuAlaProLeuAlaHisAlaAsnProProAlaProLeuAspLeuAspIleGlyIle 120  
 QY 394 TCTTACGTTTGAAGCGCTTCGCGCACTACGAGCGGTGGCGCGCGCGCGCGCGCGCGCG 453  
 DB 121 SerTyrArgPheGlyArgPheAlaAspTyrLeuSerPalaLeuAlaGlyAlaThrProAlaAla 140  
 QY 454 GTCAAGCGCGCTGTATGGTGGCGCATCAAGCTGCGCGCGCGGTATCGCGGACTTG 513

DB 141 ValAsnAlaAlaCysMetValGlyHisSerThrLeuArgAlaAlaValMetProAspLeu 160  
 QY 514 CAGCG 573  
 DB 161 GlnArgAlaAlaThrAspIleGlyIleAlaAlaMetArgAspLeuAlaGlyIleAlaMet 180  
 QY 574 GCGAGCG 633  
 DB 181 AlaSerGlyAlaIleGlyIleSerThrGlyAlaPheTyrProProAlaAlaArgAlaThr 200  
 QY 634 ACCGAGAGATCATGAGGTGTGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693  
 DB 201 ThrGluGluIleIleGlyValCysArgProLeuSerAlaHisGlyGlyIleThrAlaThr 220  
 QY 694 CACATGCG 753  
 DB 221 HisMetArgAspGlyGlyIleHisIleValAlaAlaLeuGluGluThrPheArgIleGly 240  
 QY 754 GCGAGCTGAGCTGCCGCTGTGATCTCGACACGAGCATAGGCGCGCGCGCGCGCGCGCG 813  
 DB 241 ArgGluLeuAspValProValIleSerHisIleValMetGlyGlnProAspPhe 260  
 QY 814 GCGCGCTCG 873  
 DB 261 GlyArgSerArgGluThrLeuProLeuIleGluAlaAlaMetAlaArgGlnAspValSer 280  
 QY 874 CTGAGCGCTATCCCTACGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933  
 DB 281 LeuAspAlaTyrProTyrValAlaGlySerThrMetLeuArgGlnAspArgValIleLeu 300  
 QY 934 GCGGAGAGCACATCATCATCTGCGAGACCCCTCCCGAGCTGAGCGCGCGCGCGCGCG 993  
 DB 301 AlaGlyArgThrIleIleThrTrpCysArgProPheProIleuSerGlyArgAspLeu 320  
 QY 994 GATGAAGTCCG 1053  
 DB 321 AspGluValAlaAlaGluArgGlyLysSerLysTyrAspValProGluLeuGlnPro 340  
 QY 1054 GCGCGCGCATCTACTTCTATGATGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 1113  
 DB 341 AlaGlyAlaIleTyrPheMetMetAspGluProAspValGlnArgIleLeuAlaPheGly 360  
 QY 1114 CCGACCATATGCGCTCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1173  
 DB 361 ProThrMetIleGlySerAspGlyLeuProHisAspGluArgProHisProArgLeuTrp 380  
 QY 1174 GGCACCTTCCG 1233  
 DB 381 GlyThrPheProArgValLeuGlyHisTyrAlaArgAspLeuGlyLeuPheProLeuGlu 400  
 QY 1234 ACGCGGCTATGGAATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1293  
 DB 401 ThrAlaValAlaTrpLysMetThrGlyLeuThrAlaAlaArgPheGlyLeuAlaGlyArgGly 420  
 QY 1294 CAGTGCAGCG 1353  
 DB 421 GlnLeuGlnAlaGlyTyrPheAlaAspLeuValAlaPheAspProAlaThrValAlaAsp 440  
 QY 1354 ACCGCGACCTTGAACACCTTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413  
 DB 441 ThrAlaThrPheGluHisProThrGluArgAlaAlaGlyIleHisSerValTyrValAsn 460  
 QY 1414 GCG 1473  
 DB 461 GlyAlaProValAlaTrpGlnGlnAlaAlaPheThrGlyGlnHisAlaGlyArgValLeuAla 480  
 QY 1474 CGCAGCGCGCGCG 1485  
 DB 481 ArgThrAlaAla 484

RESULT 2  
 AAY84943  
 ID AAY84943 standard; Protein; 484 AA.





QY 1048 -----CACCCGCGCGCCACATCTTATGATGAGAACCCAGCG 1092  
 Db 376 LeuAlaLysAspGlyThrAsnAlaGlyMetLeuValPheLeuMetSerGluGluAspVal 395  
 QY 1093 CAGCGCATCTCGCGCTGCGCCGACCATGATCGGCTCCGAGCGGCTCCGACGACGAG 1152  
 Db 396 GluArgIleuSerHisProTyrSerMetIleGlyThrAspGlyLeuAspSerGlyIle 415  
 QY 1153 ---CGCCCGCATCGCGCGCTGCGGCGCATCTCCCGCGGCTGCTGGGCACTATCGCGC 1209  
 Db 416 GlyLeuProHisProAlaGlyAlaTyrGlyThrPheProArgValIleGlyArgTyrValArg 435  
 QY 1210 GACCTGGCGCTGTCCTCGCTGAGACGCGGATGAGAGATGACGCGGCTGACGCGCGC 1269  
 Db 436 GluTyrLysLeuLeuArgLeuGluAspAlaIleArgLysMetThrSerLeuProAlaLeu 455  
 QY 1270 CGCTTCGCGCTGCGCGCGCGGCGGACGCTGACGCGCGGCTGCTGCGGACCTGGTGTG 1329  
 Db 456 LysLeuGlyLeuLysAspAlaTyrGlyLeuValLysGluGlyMetThrPalaAspLeuValIle 475  
 QY 1330 TTCGACCGCGCGACGCTGCGCGGATACCGCGCATCTCGAACACCTACGAGCGCGCGC 1389  
 Db 476 PheAspProHisArgValLysAspAlaIleThrTyrThrAsnProArgLeuProProAsp 495  
 QY 1390 GGCATCCATTCGCTGATGATCAACGCGCGCGCTGCGCAAGAGACGCGCTGACGCGC 1449  
 Db 496 GlyIleTyrValIleValAlaGlyValLeuSerValGluAsnGlyGluLeuThrGly 515  
 QY 1450 CACGATGCGCGCGCGCTGCTGCGACGCGCGC 1482  
 Db 516 AspAlaGlyValValIleArgArgThrSer 526  
 RESULT 4  
 AAB47456  
 ID AAB47456 standard; protein; 558 AA.  
 XX AAB47456;  
 AC  
 XX  
 DT 13-DEC-2001 (first entry)  
 XX  
 DE D-aminoacylase.  
 XX  
 KM D-aminoacylase: N-acetyl-D-tryptophan; N-acetyl-D-phenylalanine; PCR;  
 KM N-acetyl-D-valine; N-acetyl-D-leucine; N-acetyl-D-methionine; primer;  
 KM probe; synthesis; detection; D-amino acid; substrate specific; amplify;  
 KM thermal stability; polymerase chain reaction.  
 OS Hypomyces mycophilus.  
 XX  
 PN EP1120465-A1.  
 XX  
 PD 01-AUG-2001.  
 XX  
 PF 25-JAN-2001; 2001EP-0101739.  
 XX  
 PR 27-JAN-2000; 2000JP-0019080.  
 PR 22-MAY-2000; 2000JP-0150578.  
 XX  
 PA (DAIL ) DAICEL CHEM IND LTD.  
 XX  
 PI Mitsubashi K, Yamamoto H, Matsuyama A, Tokuyama S;  
 XX  
 DR WPI; 2001-551332/62.  
 DR N-PSDB; AAH43261.  
 XX  
 PT Novel D-aminoacylase-encoding gene derived from filamentous fungus  
 PT Hypomyces mycophilus, useful for producing D-tryptophan from  
 PT N-acetyl-D-tryptophan, useful as medicinal raw material -  
 XX  
 PS Claim 1; Page 17-20; 33pp; English.  
 CC This sequence shows a D-aminoacylase polypeptide. D-aminoacylase has

CC physicochemical properties that include the action of the enzyme on  
 CC N-acetyl-D-amino acids to produce the corresponding D-amino acids,  
 CC and substrate specificity, where the enzyme acts on  
 CC N-acetyl-D-tryptophan, N-acetyl-D-phenylalanine, N-acetyl-D-valine,  
 CC N-acetyl-D-leucine, and N-acetyl-D-methionine, but not on  
 CC N-acetyl-D-tryptophan, N-acetyl-D-phenylalanine, N-acetyl-D-valine,  
 CC N-acetyl-D-leucine, or N-acetyl-D-methionine. Fragments of the  
 CC D-aminoacylase cDNA are useful as primers or as probes for synthesizing  
 CC or detecting the full length cDNA. D-aminoacylase is useful for  
 CC producing D-amino acids. The enzyme has substrate specificity,  
 CC thermal stability and produces D-amino acid efficiently by incubating  
 CC the fungus derived D-aminoacylase with N-acetyl-D-amino acid under  
 CC proper conditions. The enzyme has high enzymatic activity for  
 CC N-acetyl-D-tryptophan and is excellent in industrial applicability.  
 CC The recombinant polypeptide of D-aminoacylase can be manufactured at  
 CC a low cost and in large quantities.

Sequence 558 AA;

# Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
61	612.00	558	171	76	226	13	
4	45.328	558	76	226	13		
Best Local Similarity:	31.388						
Query Match:	18.538						
DB:	22						

US-10-009-782-1 (1-1758) x AAB47456 (1-558)

QY 61 GACCTGCTGCTCGCGCGCGGCGGACCCATGATGAGAACCCGCGGCGG---CGC 117  
 Db 4 GluIleuPheHisSerAlaIleThrValIleThrGlyAspIleAlaIleGlnProPheVal 23  
 QY 118 GCGGACCTGCGGCGGCGGCGGCGGACCGCATCGCGCATCGCGATCGTGC-----GAC 171  
 Db 24 AlaAspValLeuValSerLysGlyLeuIleAlaLysIleGlyAsnProGlySerIleAsn 43  
 QY 172 GCGCGCGCGGACCGCGG---GTCGACGTGTCGCGGCTGTCGCGCGCGCGCTGATC 228  
 Db 44 AlaThrProAspThrArgHisIleAspValThrGlyTyrIleLeuSerProGlyPheIle 63  
 QY 229 GACTCGCACACCGACGACGACACTACCTGCTGACGCGCGCGCATGACGCCCAAGATC 288  
 Db 64 AspMetHisAlaHisSerAspLeuTyrLeuLeuSerHisAspAlaGlnAlaLysIle 83  
 QY 289 TCGCAGGCGGCTGACGAGGTGTACGCGGCAATTGCGGATGACGCTGCGCGCGCGGCG 348  
 Db 84 ThrGlnGlyCysThrThrGluValValGlyGlnAspGlyIleSerTyrAlaProIleArg 103  
 QY 349 CACGCGACCGCG-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372  
 Db 104 AsnValAspIleLeuArgAlaIleArgGluGlnIleAlaGlyTyrAsnGlyAsnProThr 123  
 QY 373 GAC-----CTGCTGAGCAGAGCGCGCTCTTACCGTTTTCGAGCGCTTCCG 417  
 Db 124 AspGluGluCysArgThrThrLeuLysGlyValGlyMetPheGluThrPheIleThrIleGly 143  
 QY 418 GACTACCTGAGCGGCTTGGCGGCGGCGCGCGCGCGCTGACGCGCGCGCTGATGAGGCG 477  
 Db 144 GluTyrLeuAspPheLysLeuGluArgAsnArgThrAlaThrAspValAlaMetLeuValPro 163  
 QY 478 CATTCAACGCTGCGCGCGCGCGGCTCATGCGGACTGACAGCGCGCGCGCGCGAGAA 537  
 Db 164 GlnGlyAsnLeuArgLeuLeuAlaCysGlyProTyrAspThrProAlaSerAlaGluIle 183  
 QY 538 ATCGGCGCATGCGGAGCTGCGCGGAGAGCATGCGGCGCGCGCGCATGCGGCTTTCG 597  
 Db 184 IleGlnAspGlnIleGlnLeuLeuArgGluAlaMetAlaGlnGlyValGlyMetSer 203  
 QY 598 ACCGCGCGCTTACCG 657  
 Db 204 SerGlyLeuThrTyrThrProGlyMetTyrAlaSerThrSerGluLeuAlaSerLeuGly 223

[illegible]

```

XX 23-FEB-1999      (first entry)
DX
DE Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA22.
XX
XX Human; N-methyl-D-aspartate receptor; NMDAR2C;
KW NMDA-activated cation-selective ion channel; glutamate receptor.
OS Homo sapiens.
XX
XX US5849895-A.
XX PN
XX PD
XX 15-DEC-1998.
XX
XX 20-APR-1994;    94US-0231193.
PF 20-APR-1994;    94US-0231193.
PR 20-APR-1994;    94US-0231193.
PR 20-APR-1993;    93US-0052449.
XX
XX (SIBI-) SIBIA NEUROSCIENCES INC.
PA
XX
XX Daggett LP, Lu C;
PI
XX
XX WPI: 1999-069812/06.
DR
XX N-PDB; AAV82909.
PS
XX
XX Example 3; Columns 253-262; 203pp; English.
PT
XX The present sequence represents a human N-methyl-D-aspartate (NMDA)
CC receptor subunit (NMDAR). The nucleic acid sequence does not contain
CC the 366 5' most nucleotides, by the insertion of 11 nucleotides between
CC nucleotides 1300 and 1301, nor the 15 nucleotides at positions
CC 1960-1974, nor the 1061 3' nucleotides, as set forth in AAV82889. The
CC cDNA sequence is derived from clone NMDA21. The NMDAR subunits contribute
CC to the formation of NMDA activated cation-selective ion channels. In
CC addition to being useful for the production of NMDA receptor subunit
CC proteins, the nucleic acids are also useful as probes to identify and
CC isolate nucleic acids encoding related receptor subunits. Functional
CC glutamate receptors can be assembled from several NMDA receptor subunit
CC proteins of one type (homomeric) or from combinations of subunit proteins
CC of different types (heteromeric). The present invention also comprises
CC methods for using such receptor subunits to identify and characterise
CC compounds which affect the function of such receptors , e.g. agonists,
CC antagonists and modulators of glutamate receptor function. The invention
CC also comprises methods for determining whether unknown protein(s) are
CC functional as NMDA receptor subunits.
XX
XX
SQ Sequence      1212 AA;

Alignment Scores:
Pred. No.:          2,396-21           Length:         1212
Score:              411.00             Matches:         188
Percent Similarity: 33.61%             Conservative:     59
Best Local Similarity: 25.58%            Mismatches:      230
Query Match:        12.46%              Indels:          258
DB:                  20                 Gaps:            41

US-10-009-782-1 (1-1758) x AAW87503 (1-1212)

QY 32 CCATGTCGCC-----ATCGGAT-----CCGACG 55
   |||||
Db 370 ProcsyPFOALAlAGLYArGaLarThrProSerAlaAlAGLYThrTrpProThr 389
   |||||
QY 56 CCTTGCACCTGTGGCGGGGGCCACCCTCATGAGCGGCAACAACCCCG--GAG 112
   |||||
Db 390 ProSerAlaValArgAspSerAlaSerThrSerArgSerTrpProLutrpser 409+
   |||||
QY 113 GGCGCGCG-----ACCTGGCGGCGGGGGAGCACCGCATGCCGCCCATGCCGATC 163
   ::|||
Db 410 AsnSerProThrTrpCysThrTrpProThrAlaSerMetAlaSeglyCysAlaLaIayr 429

```

```

OY 164 -----TGTGGAGCGCGCGGACACCGGGTGCAGCTGTGGGCC 205
DB 430 GLYThrAlaLeuLylarGcysThrThrSerGlyGlnThr----- 442
OY 206 TGTGGTGGCGCGCGCTTCAGACTCGACACCGACGACGACAACTACTGCTCAGGC 255
DB 443 TTPProSerAlaProSerProSerMetArgAsn----- 453
OY 266 GTGCGACATGAGCGCCAGATCTCCAGGGGCTCACACAGGTGTACGGGCAATTGCG 325
DB 454 -----AlaProArgSerThrSerLeuTyrProLeuThrPargArg 466
OY 326 GCATCAGCGCTGCGCGCTGGCGGACCGCAACCGCGCG-----CGCCCGGAGACTGTGG 382
DB 467 AlaSerValTTP-----TTPLeuAlaMetAlaProSerProProArgProSerTTP 484
OY 383 AGGAGGCGGCTTACCTTTCGAGCGCTTGGCGGACTGACGAGCGGTGGCGGCA 442
DB 485 SerHisAlaLeu-----GlnCysGlyCys 493
OY 443 CGCCGCGCGCTCAACCGCGCTGTATGTGTGGCATTCACAGCTGGCGCGCGGTCA 502
DB 494 Leu-----SerCysAlaSerLeuTTPPro-----SerProSerSer 506
OY 503 TGCCGCACTTGCAGCGCGCGCGCGACGAGGAAATCGCGGCA----- 547
DB 507 CysSerSerThrSerAlaLeuSerAlaThrThrArgThrSerProGlnAlaArgLeuSer 526
OY 548 -----TGCGGAGACTGG----- 559
DB 527 LeuSerAlaSerProCysGlyCysGlySerArgTTPSerSerThrThrGlnCysProSer 546
OY 560 -----CCGAGGAGGCAATG-----CCAGCGCGCGCATCG 589
DB 547 ArgThrArgGlyAlaProProAlaArgSerTTPThrThrSerGlyProSerLeuLeuSer 566
OY 590 GCATTTCGACCGCGCGCTTCACCGCGCG-----CGCGCGCGCGCACCAACGAGAGA 643
DB 567 Ser-----SerSerProAspThrArgProThrTTPProSerSerTysSerAsnThr 584
OY 644 TCATCGAGGT----- 655
DB 585 SerThrLeuCysArgAlaSerValThrArgSerPheSerGlyLeuLysIleSerThrHis 604
OY 656 -----GCCGCGCGCTGACGCGCGCATGGCGGATCGAC 688
DB 605 LeuSerAlaSerAlaArgCysProThrAlaAlaArgSerGlyThrSerAlaValThrThr 624
OY 689 -----CCACCCACATGCGGAGAGGACGACACATCGTGGCGCGCTGGAGAGA 739
DB 625 ValThrCysThrProThrTTPSerSerSerThrSerAlaArgTTP--ArgThrArgSer 643
OY 740 CCTTCGCGATCGCGCGCGAGCTG-----AGTGGCGG 772
DB 644 ProAlaSerArgTTPLeuSerTTPMetProSerSerMetLeuLeuSerSerThrThr 653
OY 773 TGTGATCTGCGACCAAGTATGCGGACCGCAATTTCGCGCGCTGCG----- 823
DB 664 TTPGlnAlaArgThrArgAlaAlaSerTTPSerProLeuGlyLeuAlaArgSerLeuLeu 683
OY 824 -----GCGAGAGCTGCGCG-----TGATCGAGGCGCGCATGCGCG 859
DB 684 ProLeuAlaThrAlaSerProCysArgArgThrProThrGlySerIlyProThrTTPArg 703
OY 860 GCCAGAGACTC----- 871
DB 704 SerCysSerSerTTPLeuGlyGlnArgHisArgAsnTTPArgGlnCysGlySerGlnGly 723
OY 872 -----CGGTGACGCGGTATCCCTAGTGGCGCGCGCTCCACCATGCTCAAGCAG 919
DB 724 SerAlaArgMetArgArgThrArgAlaAlaSerTTPThrThrThrTTPGlnAlaSer 743

```

```

OY 920 ACCGCTGTGTGGTGGCGGAGCGACACATCATCTGTTGCAAGCCCTTCCCGAATGA 979
DB 744 SerThrCysCysTTP----- 748
OY 980 GCGGCGCGCATGTGATGATGCGCGCGCGGCGGCAATCCAAATGACAGCTGTGCG 1039
DB 749 -----TTPProThrLylThrProCysTTPSerSerProGlySerThrTTPSer 764
OY 1040 CGG-----AGCTGAGCGCGCGCGCGCATCTCATGATGACAGACACCGCATGACG 1096
DB 765 ThrGlySerCysAlaThrArgCysPro-----ThrHisProSerTTPThr 779
OY 1097 GCATCTGTGCGTTCGCGCGCGACCATGATGCGGCTCGAGCGCG-----TGCGGACAGCAGC 1153
DB 780 SerCysTTPLeuSerAlaGlyAlaSerThrAlaAlaSerAlaGlyCysArgAlaSerPro 799
OY 1154 GCGCCCATCCGCGCGCTGTGGGACACTTCCGCGGCTGCTGGGCGCATGACGCGGAC 1213
DB 800 AlaHis-----ArgGlyArgProAlaArgThrSer--ArgProAlaArgProArg 815
OY 1214 TGCGCTGTTCGCGTGGAGAGCGGCGATGAGATGACCGGCTTACCGCGCGGCT 1273
DB 816 ProAlaCysSerAlaArgPheCysArgGln-----ProAlaThrTTPProArgArg 831
OY 1274 TGCGCTGTGCGCGGCG-----GGGCGGAGCTGC 1300
DB 832 AlaAlaThrProTTPThrAlaProLeuAlaProSerArgIleGlyValAlaAlaAlaVal 851
OY 1301 AGCGCGGCTTACTTCCGCGGACTGTGTTCGACCGCGCGCGGCGGATACCGGCA 1360
DB 852 ArgProHisArgProProProAlaArgProArgIlyLeuAlaProAlaHisAlaCysProPro 871
OY 1361 CCTTCGAACACCTTACCGAGCGCGCGCGCATTCATTCGCGTTCATACCGCGCGCG 1420
DB 872 ProThrArgProGlnSerArgAlaProArgAlaGlyAspArgGlnThrGlyValAlaArg 891
OY 1421 CGGTCTGGCAAGAGACGCGCTTACCGCGCGCATCGCGCGCGCTGCTGCGACGCGG 1480
DB 892 ArgLeuCysAlaGlyLeuArgSerProArgAlaAlaProArgThrArgIlyArgProCys 911
OY 1481 CGCGCTGAGCGCGCGCGCACCTTACATTCGCGCGTGAAGGCGGCGGCTGCGCGCG 1540
DB 912 ProThrSerProLylCysArgAlaAlaGlnPro-----GlyArgArgGlyGlyArg 928
OY 1541 TCCCAACCTTGAAGCAACCGCTACATGAGCCCTCCCTCCGTCGACATGACGCGCG 1600
DB 929 CysGlyProGlyThrAlaGlyThrSerArgProProSerGlyProCysArgProArg 948
OY 1600 ----- 1600
DB 949 AlaValThrThrAlaProPheLeuGlnProThrAspProAlaAlaProSerSerArgSer 968
OY 1601 -----CCGATATCTGTGGCAAGAGATGATGCGCGCGCTGCGCGCGCGCGCA 1651
DB 969 SerArgSerProArgSerTTP-----ArgPheCysArgCysSerVal 982
OY 1652 AGG-----CCGGAATAATGACCTTGCAGACCTTGCAGCGCGCGCG-----GCATCG 1699
DB 983 ArgSerSerTTPProGlyGlyArgPro-----CysThrArgProGlyProGlyAla 999
OY 1700 CGGTCTGCA-----CCGTGCAAGCGCGAGCTGGGCGCA 1735
DB 1000 ArgAlaArgValThrLeuProCysProAlaPro--TTPProArg 1013

```

RESULT 6  
 AA#87504  
 ID AA#87504 standard; Protein; 1061 AA.  
 XX  
 AC AA#87504;  
 XX  
 DT 23-FEB-1999 (first entry)  
 XX  
 DE Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA24.



The present sequence represents a human N-methyl-D-aspartate (NMDA) receptor subunit (NMDAR). The nucleic acid sequence does not contain the 860 5'-most nucleotides. The nucleic acid sequence does not contain the 860 5'-most nucleotides, has an additional 11 nucleotides (AAV82891) between nucleotides 1300 and 1301, an additional 24 nucleotides (AAV82890) inserted between nucleotides 23050 and 2351, as set forth in AAV82889. The cDNA sequence is derived from clone NMDA24. The NMDAR subunits contribute to the formation of NMDA-activated cation-selective ion channels. In addition to being useful for the production of NMDA receptor subunit proteins, the nucleic acids are also useful as probes to identify and isolate receptors can be assembled from several NMDA functional glutamate receptors can be assembled from several NMDA receptor subunit proteins of one type (homomeric) or from combinations of subunit proteins of different types (heteromeric). The present invention also comprises methods for using such receptor subunits to identify and characterize compounds which affect the function of such receptors, e.g., agonists, antagonists and modulators of glutamate receptor function. The invention also comprises methods for determining whether unknown protein(s) are functional as NMDA receptor subunits.

US-10-009-782-1 (1-1758) x AAW87504 (1-1061)

QY	206	TGTTGGTGGCGGGCGGCTTCAACGACATGCGACACCCAGACAGACAACTACTGCTCAGGC	265
Db	279	TyrProSerAlaProSerProSerMetAlaGln	285
QY	266	GTCCGACATGACGCCCAAGATCTGCGAGGGCGTCCACCAGGTGGTCAAGGCAATTGCG	325
Db	290	-----AlaProArgSerThrSerLeuTyrProLeuThrPArg	305
QY	326	GCATCAGCGCTGGCGCGCTGGCGGCACGCCAACCCGCCG	385
Db	303	AlaSerValTyr-----TyrLeuAlaMetAlaProSerProProArgProSerTyr	320
QY	383	ACGAGAGCGGCTTACCGTTCGAGAGCGCTTCCGACTCTGAGCGGTTGGGGCCA	445
Db	321	SerHis1LeuAlaLeu-----GlnCysGlyCys	325
QY	443	CGCGCGGCGGCGCTCAACACCGCGCTGTATGTTGGCCATTCAACGCTGGCGCGCGCTCA	505
Db	330	Leu-----SerCysAlaSerLeuTyrTrpPro-----SerProSer	345
QY	503	TGCGGACTTGCAGCGCGCGCGCACCGACGAGAAATGGCGCA	545
Db	343	CysSerSerThrSerAlaLeuSerAlaThrThrArgTrpSerProGluAlaArgSerPro	365
QY	548	-----TGCGGGACCTGG-----	555
Db	363	GlyAlaGlnLeuSerLeuSerAlaSerProCysGlyCysGlyArgTyrPserSerThr	385
QY	560	-----CCGAGAGAACCATGG-----	575
Db	383	ThrGlnCysProSerArgThrArgGlyAlaProProAlaLeuSerThrPheTyrPserGly	405
QY	575	CCAGCGGCGGCATCGGCATTTCGACCGCGCGCTTCAACCGCGG	625
Db	403	ProSerLeuLeuSerSer-----SerSerProAspThrArgProThrTrpProProSer	420
QY	629	CCACCACGAGAGATCATCGAGTGT	655
Db	421	SerTyrSerAsnThrSerThrLeuCysArgAlaSerValThrArgSerPheSerGlyLeu	440
QY	656	-----CCGCGCGCGCTAGCGCGG	673
Db	441	LysIleSerThrHisLeuSerAlaSerAlaGlySerProThrAlaAlaArgSerGlyThr	460
QY	674	ATGGCGGATATACGCCACCCACATGCGGCAGAGGCGACATCGTGGCGCGCTGG	733
Db	461	SerAlaValThrThrValThrCysThrProThrTyrPserSerSerThrSerAlaArgTyr	480
QY	734	AGG-----AAACCTTCGCGATCGGCGCGCGAGCTGG	763
Db	481	ArgThrArgSerProAlaSerArgTyrPAlaLeuArgLeuSerLeuSerProCylSerTyr	500
QY	764	-----ACGTGCGGCTGGTGTATCTGCACCAAGAGTCA	796
Db	501	MetProSerSerMetMetLeuLeuSerSerThrThrTyrGlnAlaArgThrArgAlaAla	520
QY	797	TGGGCGACGCCCATTTGCGCGCGTGGC	835
Db	521	SerTyrSerProLeuGlyLeuAlaAlaArgSerLeuLeuProLeuAlaThrAlaSerProCys	540
QY	836	CGC-----TGATGAGCGCGCGCATGCGCGCGCGAGACGCT	871
Db	541	ArgThrThrProThrGlySerGlyProThrTyrPargSerCysSerSerTyrPglyThrGlu	560
QY	872	-----CGCTGGACGCGT	883
Db	561	ArgHisArgAsnTyrPArgGlnCysGlySerGlnGlySerAlaArgMetArgThrArg	580
QY	884	ATCCCTACGTGGCGCGCTCCACCATGCTCAAGACAGACGCGGTGCTGGCGGAGCGCA	943
Db	581	AlaAlaSerTyrPheSerThrThrTyrPglAlaSerSerThrCysCysTyr-----	597

```

OY 944 CCATCATCACTGGTCCAGAGCCCTCCCGACAGCGGGCGACCTGGATGAGTGC 1003
DB 598 -----TrrProTrrPely 601
OY 1004 CGCGCCAGCGCGCAATCCAGTACAGCTGTGCCG---AGCTGCAGCGCGCGCG 1060
DB 602 TrrProCysTrrSerProGlySerThrTrrSerThrCysSalatThrArgCys 621
OY 1061 CCATCACTTCAATGATGAGCAACCGAGCTGACGCGCATCCCTGGCTGGCCCGCA 1120
DB 622 Pro-----ThrHisProSerTrrTrrSerTrrSerTrrPleuSerAlaGlyAla 636
OY 1121 TGAATCGCTCCGACGCGC---TGCCGACAGCAGCGCGCGCATCCGCGCTGGGCGCA 1177
DB 637 SerThrAlaAlaSerAlaGlyCysArgAlaSerProAlaHis-----ArgGlyArg 653
OY 1178 CCTCCCGCGGGGTCTGGGGCAGCTATCGCGCGCAGCTGGCGCTGTCCCTGGAGAGG 1217
DB 654 ProAlaArgThrSer---ArgProAlaArgProArgProAlaCysSerArgPheCysArg 672
OY 1238 CGGTATGAGAAATGACCGGCTACCGCGCGCTGGCGCTGGCGCGCGC----- 1288
DB 673 Gln-----ProAlaThrTrrProArgArgAlaAlaThrProTrrPrrAlaPro 688
OY 1289 -----GCGGCGAGCTGCAGCGCGCGGTACTTCCGCGACCTGG 1324
DB 689 LeuAlaProSerArgIleGlyValAlaAlaAlaValArgProHisArgProProAlaArg 708
OY 1325 TGGTGTCCACCGCGCGCGAGGTGGCGGATACCGCCACCTTCGAAACCTACCGAGCGC 1384
DB 709 ProArgGlyLeuAlaProAlaHisAlaCysProProProThrArgProGlnSerArgAla 728
OY 1385 CGCGCGCATTCATTCGATCCGTACGTACAGCGCGCGCTGCTGCAAGACAGCGCTCA 1444
DB 729 ProArgAlaGlyAspArgGlnThrGlyValAlaArgArgLeuGlyAlaGlyLeuArgSer 748
OY 1445 CGCGCCAGCATCGCGCGCGCGCTGCTGCAGCGCGCGCTGAGCGCGCGCGCGCGCT 1504
DB 749 ProArgAlaAlaProArgArgArgArgGlyArgProCysProThrSerProGlyCysArgAla 768
OY 1505 TACATCCGCGCTGAAACGGGGCGCGCTGCCCGCCCTCCACCCCTGGAGCGCAACCGCT 1564
DB 769 AlaGlnPro-----GlyArgArgGlyGlyArgCysGlyProGlyThrAlaGlyGly 785
OY 1565 ACATGGCCCTCCCTCGCTCGCAATACGCGCGCAC----- 1600
DB 786 ThrSerArgProProSerGlyProCysArgProArgAlaValThrAlaProPheLeu 805
OY 1601 -----CCGATATCGTGGCA 1615
DB 806 GluProThrAspProAlaAlaAlaProSerSerArgSerSerArgSerProArgSerTrrP--- 824
OY 1616 AGGAAGTATGGCGCGCGCTGCGCGCGCGAGCGAAG-----CCCGGAAATGA 1666
DB 825 -----ArgThrCysArgCysSerValArgSerSerTrrProGlyGlyArg 839
OY 1667 CCGTCGAGACCTGTTCGAGCGCGCAGG---GCATCGCGGTCTGA-----CCCTGT 1714
DB 840 Pro-----CysThrArgProGlyProGlyAlaArgAlaArgValThrLeuProCys 856
OY 1715 CCAGGCGCGAGCTGGCGCA 1735
DB 857 ProAlaPro---TrrProArg 862

```

```

XX Human: N-methyl-D-aspartate receptor; NMDA; NMDAR1A; Ionotropic;
KW glutamate receptor; drug screening; animal model; disease diagnosis;
KW genetic screening.
XX Homo sapiens.
FH Key Location/Qualifiers
FT Misc-difference 280
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 304
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 326
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 347
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 422
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 438
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 464
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 477
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 501..502
FT /label- unknown
FT /note- "encoded by TGATGA"
FT Misc-difference 590
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 712
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 744..745
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 843
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 848
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 1009
FT /label- unknown
FT /note- "encoded by TAA"
FT Misc-difference 1152
FT /label- unknown
FT /note- "encoded by TGA"
FT Misc-difference 1209
FT /label- unknown
FT /note- "encoded by TAG"
FT Misc-difference 1227
FT /label- unknown
FT /note- "encoded by TGA"
FT /label- UNKNOWN
FT /note- "the coding sequence of this protein contains a
number of in-frame stop codons which are represented
by Xaa in the protein sequence"

```

```

XX US6111091-A.
XX 29-AUG-2000.
XX 29-SEP-1997; 97US-0940086.
XX 20-APR-1994; 94US-0231193.

```

PR 20-APR-1993; 93US-0052449.  
 PA (MERI) MERCK & CO INC.  
 XX  
 PI Daggett LP, Lu C;  
 XX  
 DR WPI: 2000-578607/54.  
 XX N-PSDB: AAA95032.

PT Novel DNA fragment encoding human N-methyl-D-aspartate receptor subunit  
 PT for identifying mutations and for developing drugs against various  
 PT disease states  
 XX

Example 3; column 255-264; 205pp; English.

The present sequence is a subunit (designated NMDAR1) of the human  
 CC N-methyl-D-aspartate (NMDA) receptor. This is an ionotropic glutamate  
 CC receptor which contains cation-specific ligand-gated ion channels. The  
 CC protein and its coding sequence can be used in disease diagnosis and in  
 CC research to identify other similar proteins. They can also be used as  
 CC probes, for example in genetic screening, and in drug screening, as well  
 CC as enabling the production of animal disease models.

SO Sequence 1232 AA;

Alignment Scores:  
 Pred. NO.: 5,32e-19 Length: 1232  
 Score: 378.50 Matches: 189  
 Percent Similarity: 33.15% Conservative: 58  
 Best Local Similarity: 25.37% Mismatches: 233  
 Query Match: 11.47% Indels: 265  
 Gaps: 21 44

US-10-009-782-1 (1-1758) x AAB26239 (1-1232)

```

OY 32 CCAATGTCCTC-----AATCCGAT-----CCAGC 55
DB 374 ProcySerProAlaIaIaGlyArgAlaThrThrProSerAlaIaIaGlyThrProProThr 393
OY 56 CATTGACCTGCTGCTGCGCGCGCGCCCTCATCGAGCGCAGCAACCCCGG---GGC 112
DB 394 ProSerSerAlaValArgAspSerAlaSerThrSerSerArgSerTrpProGluTrpSer 413
OY 113 GCGCGCCG-----ACCTGG-----GGTGGCGGCGC 139
DB 414 AsnSerProThrThrCysThrTrp**ProThrAlaSerMetAlaSerGlyCysAlaIaIa 433
OY 140 ACCGATCGCCCGCATCGCGCATCTGTCGAGCGCGCGCGACACCCGGTCGACGT 199
DB 434 TyrGlyThrAla**LeuGlyArgCysThrThrSerGlyGlnThr----- 448
OY 200 CGGCGCTGTGTGCGCGCGCGCGCTCATCGACTCGC---ACACCGACGACGACACTAC 256
DB 449 -----TTPProSerAlaProSerProSerMetArgAsnAlaProArgSer**ThrSer 466
OY 257 TGTCTAGGCGTGTGCGACATGACGACCAAGATCTGCGACGGGTCACACGGTGTACAG 316
DB 467 Leu-----TyrProLeuThrProArgAlaSerVal**Trp----- 478
OY 317 GCAATGTGGGATCAGCCTGCGCGCGCGCGCGCAACCCCGCG---CCGCCCTGG 373
DB 479 -----TTPLeuAlaMetAlaProSerProProArg 489
OY 374 ACTGTGTGAGAGAGGGGCTTACCGTTTCAGAGCGCTTCGCCACTACCTGAGCGGT 433
DB 490 ProSerTrpSerAlaIaIaLeu-----Gln 498
OY 434 TGGGCGCCAGCGCGCGCGCGCGCTATGCTATGCTGAGGCGCATTCACGCTGGCG 493
DB 499 CysGly-----CysLeuSerCysAlaSerLeuTrpTrpPro-----Ser 513
OY 494 CCGGCGTCATCGCGGACTTCACGCGCGCGCGCGACGACGAGGAATCGCGGCGCA----- 547
  
```

```

DB 514 ProSerSerCysSerSerThrSerAlaLeuSerAlaThrThrArgThrSerProGluAla 533
OY 548 -----TGGCGGACCTGG----- 559
DB 534 ArgLeuSerLeuSerAlaSerProCysGlyCysGlyArgTrpSerSerThrThrGln 553
OY 560 -----CCAGGAGACCATG-----CCAGC 580
DB 554 CysProSerArgThrArgGlyAlaProProAlaArgSerThrThrTrpSerGlyProSer 573
OY 581 GCGCATCGGATTCGACCGCGCGCTTACCGCGCG---CCGCCCGCGCACCA 634
DB 574 LeuLeuSerSer-----SerSerProAspThrArgProThrTrpProProSer***Ser 591
OY 635 CCGAGAGAG----- 643
DB 592 LysSerAsnThrSerThrLeuCysArgAlaSerValThrArgSerPheSerGlyLeuLys 611
OY 644 -----TCATGAGAGCTT-----GCCGCGCGCTGACCGCGCATG 676
DB 612 IleSerThrAlaLeuSerAlaSerAlaArgCysProThrAlaAlaArgSerGlyThrSer 631
OY 677 GCGGATCTAG-----CCACCATGCGCGAGAGAGGAGGACATCGGCGC 727
DB 632 AlaValThrThrValThrCysThrProThrTrpSerSerSerThrSerAlaArgTrp--- 650
OY 728 CGTGTGAGAAACCTTCGCGCATCGCGCGCGCGCGCTG----- 763
DB 651 ArgThrArgSerProAlaSerArgTrpLysTrpMetProSerSerMetLeuLeu 670
OY 764 ---ACGTGCGGTGTGTCTGTCGACCAAGATGATGGCGACCGCATTCGCGCGCT 820
DB 671 SerSerThrThrTrpGlnAlaArgThrArgAlaAlaSerTrpSerProLeuGlyLeuAla 690
OY 821 CGC-----CGGAGACGCTGCGCGCTGATGAGCGCGCA----- 853
DB 691 ArgSerLeuLeuProLeuAlaThrAlaSerProCysArg-----AlaThrProThrGly 708
OY 854 -----TGGCGCGCGAGACGCT----- 871
DB 709 SerGlyPro**ThrTrpArgSerCysSerSerTrpLysThrGluArgAsnTrp 728
OY 872 -----CGTGAGCGGT-----ATCCTACG 892
DB 729 ArgGlnCysGlySerGlnGlySerAlaArgMetAlaArgThrArg**AlaAlaSer 748
OY 893 TGGCGCGCTCCACCATGCTCAACAGACGCGCGCTGCTGGCGCGGACGACCATCATCA 952
DB 749 TrpThrSerThrThrTrpLysAlaSerSerThrCysCysTrp----- 762
OY 953 CTTGTGCAAGCCCTTCCCGCAACTGAGGCGCGGACCTGATGATGATGCGCGCGAGC 1012
DB 763 -----TTPProTrpLysTrpProCys 769
OY 1013 GCGGAAATCCAAATGACAGCTGTGTCCCG---ACGTGAGCGCGCGCGCGCATCTACT 1069
DB 770 TrpSerSerProGlySerThrTrpSerThrGlySerCysAlaThrArgCysPro----- 787
OY 1070 TCATGATGAGAAACCGACCTGACGCGCATCTGCGCGCTGCGCGCGACATGATCGGCT 1129
DB 788 -----ThrAlaProSerTrpThrSerCysTrpLeuSerAlaGlyAlaSerThrAla 804
OY 1130 CCGAGCGCC---TCCGAGAGAGAGCGCGCGCATCCGCGCTGTGGGCGACCTTCCCG 1186
DB 805 AlaSerAlaGlyCysAlaAlaSerProAlaIaIa-----ArgGlyArgProAlaArg 821
OY 1187 GGGGCTGGGCACTATGCGCGCGCGCGCTGCTGCGCGTGGAGAGACCGCGGTATGGA 1246
DB 822 ThrSer---ArgProAlaArgProArgProAlaCysSerArgPheCysArgGln----- 838
OY 1247 AGATGACCGCGC---TACCGCGCGCGCTTGGCGTGGCGCGCGC----- 1288
DB 839 -----ProAlaThrTrp**ProArgArgAla**AlaThrProTrpThrAlaProLeu 856
  
```

```

OY 1289 -----GCGGGCAGCTGACGCGCGGCTACTTCCGCCGACTGCTGG 1327
Db 857 AlaProserArgIleGlyValAlaAlaAlaValAlaArgProHlaArgProAlaArgPro 876
OY 1328 TGTTCGACCGCGCCACGGGTGGCCGACCTTCGAAACACCTTACCAGCGCGCG 1387
Db 877 ArgIleValAlaProAlaHlaAlaIlaCysProProHlaArgProGlnSerArgAlaPro 896
OY 1388 CCGGCAATCCATTCGCTGACCTCAACGCGCGCGCTGCGCAAGCAGCGCTTACCG 1447
Db 897 ArgAlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIla 916
OY 1448 GCGCAGCATGCGCGCGCTGCTCCGACGCGCGCGCTGAGCGCGCGCGCGCGCTTAC 1507
Db 917 ArgAlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIla 936
OY 1508 AATCGCGCGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1567
Db 937 GlnPro-----GlyArgArgGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGly 953
OY 1568 TGGCCCTCCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1600
Db 954 SerArgProProSerArgIleProCysArgProAlaValAlaIlaIlaIlaIlaIlaIla 1618
OY 1601 -----CCGATATGCTGGGCAAG 1618
Db 974 ProHlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIla 991
OY 1619 AATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1669
Db 992 -----ArgHlaCysArgCysSerValAlaArgSerSerTrpProGlyGlyArgPro 1007
OY 1670 TGCAGACCTGTCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1720
Db 1008 Cys**ThrArgProGlyProGlyAla--ArgAlaIlaIlaIlaIlaIlaIlaIlaIla 1026
OY 1721 CCGAGCTGGGCGCGCA 1735
Db 1027 Pro--TrpProArg 1030

RESULT 8
AA47961
ID AA47961 standard; Protein; 1232 AA.
XX
AC AA47961;
XX
DI 05-MAR-2002 (first entry)
XX
DE Human NMDA receptor subunit SEQ ID NO 43.
XX
KW Human; N-methyl D-aspartate receptor; NMDAR; genetic screening;
XX glutamate receptor; drug; modulator; receptor.
XX
OS Homo sapiens.
XX
PN US6316611-B1.
XX
PD 13-NOV-2001.
XX
PF 29-SEP-1997; 97US-0940035.
XX
PR 20-APR-1994; 94US-0221193.
XX 20-APR-1993; 93US-0052449.
XX
PA (MERI ) MERCK & CO INC.
XX
PI Daggett LP, Lu C;
XX
WP1; 2002-096550/13.
XX
DR N-PSDB; AB199189.
XX
PT Novel nucleic acid encoding human N-methyl D-aspartate receptor 2A

```

PT subunit, useful for expressing the subunit and as probes in genetic screening techniques, to identify patients having neuropathological conditions

Example 3; Column 257-264; 206pp; English.

The invention relates to nucleic acids (I, AA199167-AA199196, ABN05910) encoding human N-methyl D-aspartate receptor (NMDAR) subunits. The nucleic acids are useful for the production of NMDA receptor subunit proteins (ABN56503-ABN56530, AA47961), as probes and to identify and isolate nucleic acids encoding related receptor subunits. Genetic screening can be carried out using the nucleotide sequences as probes to determine if any abnormalities exist with respect to any of the endogenous glutamate receptors. Also, patients having a family history of disease states related to glutamate receptor dysfunction can be screened to determine if they are also predisposed to such disease states. Expression vectors comprising (1), receptor subunits encoded by (1) and host cells comprising (1) are useful for producing selected NMDAR subunits and in specific combinations, as well as antibodies to the receptor subunits substantially free of contamination from many other NMDAR subtypes. The availability of desired receptor subtypes allows the observation of the effect of a drug substance on a particular receptor subtype or combination of NMDAR subunits and to perform initial in vitro screening of drugs or modulators in a test system that is specific for humans, a human NMDAR subtype or the combination of NMDAR subunits.

Sequence 1232 AA;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
5,32e-19	378.50	1232	189	58	233	265	44
Percent Similarity:	33.15%						
Best Local Similarity:	25.37%						
Query Match:	11.47%						

US-10-009-782-1 (1-1758) x AA47961 (1-1232)

```

OY 32 CCATGTCC-----AATCGATP-----CCGACG 55
Db 374 ProCysProAlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIla 393
OY 56 CCTCGACCTGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 412
Db 394 ProSerSerAlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIla 413
OY 113 GCGCGCGCG-----ACCTG-----GCGTGGCGGCG 139
Db 414 AsnSerProHlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIla 433
OY 140 ACCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
Db 434 TyrGlyThrAla**LeuGlyArgCysThrThrSerGlyGlnThr----- 448
OY 200 CGGCGCTGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 448
Db 449 -----TrpProSerAlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIla 466
OY 257 TGTCTAGGCGTGGCAGCATGACGCCAGATCTGCGAGCGCGTCAACGAGTGTCAAG 316
Db 467 Leu-----TyrProLeuThrPargAlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIla 478
OY 317 GCAATTTGGGCGCATGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373
Db 479 -----TyrLeuAlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIla 489
OY 374 ACGTCTGAGAGAGCGCGCTTACGCTTTCGAGCGCGTTCGCGACTACCTGAGCGGT 433
Db 490 ProSerTrpSerHlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIlaIla 498
OY 434 TGGCGGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493
Db 434 TGGCGGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493

```

Oy	1247	AAGTACCGGCC-----TGACCGGCCGCGCTTCGCCTGGCGCGGCC-----	128
Dd	839	 ProAlaThrTrp***ProArgArgLa***AAtThrProThrPhnAlaProLeu	856
Oy	1289	-----GCGGGCAGCTGCAGGCCCGGGGTACTTCGCCAGACTGGATG	132
Dd	857	AlaProSerArGlIleGlyValAlaAlaAlaValaLarArgHIsArgProFroAlaLeuPro	876
Oy	1328	TGTTGCAGCCGCGCACAGGTGGCCCATATCCAGCACTTCGAACAACCTTACCGAGCGCGG	138
Dd	877	ArgGlyLeuAlaProAlaHisAlaCyProProThrArgProInsErnArgAlaPro	896
Oy	1388	CCGGCATCCATTCCGCTACTACTCAACGCGCGCGGCTGTGGCAAAGACAGCGGTTCAACG	144
Dd	897	ArgAlaGlaYspArpArgInthrnglValAlaLarArgLeucysAlaGlyLeuArgserrPro	916
Oy	1448	GCCAGCATGGCCGGCGCGCTGTCTGCACAGCAGCGCGCCCTGAGCCCGGCGCACCCCTTAC	1507
Dd	917	ArgAlaAlaProlArgArghArgglYArgrProcysSproThrserProglucylsrArgAlaLa	936
Oy	1508	AATCGGCGCTGAAACGGGGCGCGGTGCGCCCCCTCCAACACTGGACGAAACCGCTTCA	1567
Dd	937	GlnPro-----GlyArghArgglYArgGrCysglYProglYThrAlaGlyelYthr	953
Oy	1568	TGGCCCTTCCTCCGCTCGCATATACGGCCCAC-----	1600
Dd	954	SerArgProProSereglyProCysArgProAlaGalValThrThrAlaProPheLeuglu	973
Oy	1601	-----CCGATACTCTGGGCAGAAG	1618
Dd	974	ProThrAspProAlaAlaProSerSerArgserSerArgSerProArgsertip-----	991
Oy	1619	AAGTAGTGGCGCGCGCTCGCGCGCGCGCAAG-----CCCGGAATAATGACCC	1669
Dd	992	-----ArgThnCysArgCysSerValArghserSerTrpProglYglYArgpro	1007
Oy	1670	TGCAGACCTCTGCAGAGCCAGCGCATTCGGGTCTCGA-----CCCTTCCAGAAG	1720
Dd	1008	Cys**ThrArgProglYProglYala---ArgAlaGlyValThrLeuProCysProAla	1026
Oy	1721	CCGAGCGGGCGGCACA	1735
Dd	1027	Pro--TrpProarg	1030
RESULT 9			
AAB26240			
ID	AAB26240	standard; Protein; 1081 AA.	
XX	AA26240;		
AC			
XX			
DT	23-FEB-2001	(first entry)	
DE			
XX			
XX			
Human N-methyl-D-aspartate receptor subunit NMDAR1 #18.			
KM	Human; N-methyl-D-aspartate receptor; NMDA; NMDAR1; Ionotropic;		
KW	glutamate receptor; drug screening; animal model; disease diagnosis;		
genetic screening.			
OS	Homo sapiens.		
FH			
Key	Location/Qualifiers		
Misc-difference 116	/label= unknown		
Misc-difference 140	/note= "encoded by TGA"		
Misc-difference 162	/label= unknown		
Misc-difference 183	/note= "encoded by TGA"		
Misc-difference 183	/note= "encoded by TGA"		
/Label= unknown			

FT /note- "encoded by TGA"  
 FT Misc-difference 258  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 274  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 300  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 313  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 337..338  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 431  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 561  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 593..594  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 692  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 697  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 858  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 1001  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 1058  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT Misc-difference 1076  
 FT /label- unknown  
 FT /note- "encoded by TGA"  
 FT /label- UNKNOWN  
 FT /note- "the coding sequence of this protein contains a  
 FT number of in-frame stop codons which are represented  
 FT by Xaa in the protein sequence"  
 FT  
 FT US6111091-A.  
 FT  
 FT 29-AUG-2000.  
 FT  
 FT 29-SEP-1997; 97US-0940086.  
 FT  
 FT 20-APR-1994; 94US-0231193.  
 FT 20-APR-1993; 93US-0052449.  
 FT  
 FT (MERI ) MERCK & CO INC.  
 FT  
 FT Daggett LP, Lu C;  
 FT  
 FT WPI; 2000-578607/54.  
 FT  
 FT N-PSDB; AAA95033.  
 FT  
 FT Novel DNA fragment encoding human N-methyl-D-aspartate receptor subunit  
 FT for identifying mutations and for developing drugs against various  
 FT disease states  
 FT  
 FT Example 3; column 633-672; 205pp; English.  
 FT  
 FT The present sequence is a subunit (designated NMDAR1A) of the human  
 FT N-methyl-D-aspartate (NMDA) receptor. This is an ionotropic glutamate  
 FT receptor which contains cation-specific ligand-gated ion channels. The  
 FT protein and its coding sequence can be used in disease diagnosis and in

CC research to identify other, similar proteins. They can also be used as  
 CC probes, for example in genetic screening, and in drug screening, as well  
 CC as enabling the production of animal disease models.  
 XX  
 SO Sequence 1081 AA;  
 Alignment Scores:  
 Pred. No.: 2,11e-18 Length: 1081  
 Score: 370.00 Matches: 189  
 Percent Similarity: 32.638 Conservative: 58  
 Best Local Similarity: 24.978 Mismatches: 234  
 Query Match: 11.228 Indels: 276  
 DB: 21 Gaps: 43  
 US-10-009-782-1 (1-1758) x AAB26240 (1-1081)  
 QY 32 CCATGTCCTCC-----AATCGATT-----CCGACC 55  
 Db 210 ProCysProAlaAlaGlyArgAlaThrThrProSerAlaAlaGlyThrThrProProthr 229  
 QY 56 CATTGACATGCTGTCGGGGGGGACCCCTCATTCAGCGCAGCAACACCCGG---GGC 112  
 Db 230 ProSerSerAlaValArgAspSerAlaSerThrSerSerArgSerThrProGluThrSer 249  
 QY 113 GCGCGCGCG---ACCTGG-----GCGTGGCGGGGG 139  
 Db 250 AsnSerProThrThrCysThrTrp\*\*\*ProThrAlaSerMetAlaSerGlyCysAlaAla 269  
 QY 140 ACCGCATCGCGCCGACATGCGGATGTCGCGACCGCGCGACACCGGGTGCAGCTGT 199  
 Db 270 TyrGlyThrAla\*\*\*LeuAlaYargCysThrThrSerGlyGlnThr----- 284  
 QY 200 CGGGCTGTGTGTCGGCGCGCGCTCATGACTGCG---ACACCCAGCAGCAACTACC 256  
 Db 285 -----TriProSerAlaAlaProSerProSerMetArgAsnAlaProArgSer\*\*\*ThSer 302  
 QY 257 TGTCTAGCGGTGCCACATGACGCCCAAGATCTCGAGGGGTCTCACCGGTGTCTACGG 316  
 Db 303 Leu-----TyrProLeuThrArgArgAlaSerVal\*\*\*Trp----- 314  
 QY 317 GCANATGCGGCATCAGCTGTGGCGCGCGCGGCGACCCCAACCCCGCG---CCCCCTGG 373  
 Db 315 -----TriLeuAlaAlaMetAlaProSerProProArg 325  
 QY 374 ACCTCTGAGAGAGCGGCTCTTACCGTTTCGAGCGCTTGGCGGACTACCTGAGCGCT 433  
 Db 326 ProSerThrSerHisIleAlaLeu-----Gln 334  
 QY 434 TGGGGGACACCGCGCGCGCGCTCAACCGCGCTGTATGTGGCCATCAACGCTGGCG 493  
 Db 335 CysGly-----CysLeuSerCysAlaSerLeuTrpTrpPro-----Ser 349  
 QY 494 CCGCGGTCAATCCGACTTCGAGCGCGCGCGCGCACGAGGAAATGGCGCCA----- 547  
 Db 350 ProSerSerCysSerSerThrSerAlaLeuSerAlaThrThrArgThrSerProGluAla 369  
 QY 548 -----TGGGGGACTCG 559  
 Db 370 ArgSerProGlyAlaGlnLeuSerLeuSerAlaSerProCysGlyCysGlyArgTrp 389  
 QY 560 -----CCGAGAAACCATGG--- 574  
 Db 390 SerSerThrThrGlnCysProSerArgThrArgGlyAlaProProAlaArgSerTrpPhe 409  
 QY 575 -----CCAGCGGGCGCATCGCATTCGACCGCGCTTCTACCGCGCG--- 619  
 Db 410 TrpSerGlyProSerLeuLeuSerSer-----SerSerProAspThrArgProThrTrp 427  
 QY 620 CCGCGCGCGCGCACCGAGAGAGA----- 643  
 Db 428 ProProSer\*\*\*SerIysSerAsnThrSerThrLeuCysArgAlaSerValThrArgSer 447  
 QY 644 -----TCATCAGGTGT-----GCCGCG 661



```

Db 448 pneserGlyLeuLysIleSerThrHisLeuSerAlaSerAlaGlyProThrAla 467
QY 662 CGGTAGAGGGCCATGGCGCATCTACAGCCACATGCGAGGAGAGACATCG 721
Db 468 ArgSerGlyThrSerAlaValThrValThrCysThrProThrTrpSerSerThr 487
QY 722 TGGCGCGCGTGGAGG-----AAACCTTCGCGATCG 751
Db 488 SerAlaGlyThrArgThrArgSerProAlaSerArgThrAlaLeuArgLeuSerLeuSer 507
QY 752 GCGCGAGCTGG-----ACGTCGCGGTGATCTCGC 784
Db 508 ProGlySerTrpMetProSerSerMetMetLeuLeuSerSerThrThrTrpGlnAlaArg 527
QY 785 ACCACAGGTCATGGGCGCCATTCGCGCGCTCGC----- 823
Db 528 ThrArgAlaAlaSerTrpSerProLeuGlyLeuAlaArgSerLeuLeuAlaThr 547
QY 824 GCGAGACGCTGCGCGCATGATCGAGCGCGCA-----TGGCGCGCG 862
Db 548 AlaSerProCysArg-----ArgThrProThrGlySerGlyPro***ThrTrpArgSer 565
QY 863 AGGAGCTG----- 871
Db 566 CysSerSerTrpGlyThrGlnArgHisArgAsnTrpArgGlnCysGlySerGlnGlySer 585
QY 872 -----CGCTGAGACGCT-----ATCCTACGTCGCGCGCTCCACATGCTCAAGC 916
Db 586 AlaArgMetArgArgThrArg****AlaAlaSerTrpThrSerThrThrTrpGlnAla 605
QY 917 AGGACCGCGTGTCTGTGCGCGAGCAGCACCATCATCATCTGTGCAAGCCCTCCCGCAAC 976
Db 606 SerSerThrCysCysTrp----- 611
QY 977 TGAGCGGCGCGACCTGTATGAATCCGCGCGCGCAATTCACAGTACAGCTGG 1036
Db 612 -----TTPPTPTGlyTrpProCysTrpSerSerProGlySerThrTrp 626
QY 1037 TGCCCG---ACGTGACAGCGCGCGCGCATCTACTCATGATGAGCAACCGACGCTGC 1093
Db 627 SerThrGlySerCysAlaThrArgCysPro-----ThrHisProSerTrp 641
QY 1094 AGGCACTCCTGGCGTGGCGCGCCAGCATGATGCTGCTCCAGCGCC---TGCGCGAGAGC 1150
Db 642 ThrSerCysTrpLeuSerAlaGlyAlaSerThrAlaAlaSerAlaGlyCysArgAlaSer 661
QY 1151 AGGCGCGCATCCGCGCTGTGGGCACTTCCCGCGGCTGTGGGCACTATGCGCGCG 1210
Db 662 ProAlaHis-----ArgGlyArgProAlaArgThrSer---ArgProAlaArgPro 677
QY 1211 ACCTGGCGCTGTCCGCTGAGAGACGCGGTATGAAATGACCGCGC-----TGACCG 1264
Db 678 ArgProAlaCysSerArgThcCysArgGln-----ProAlaThrTrp***Pro 693
QY 1265 CCGCGCGCTTGGCGCGCGCGC-----GCG 1291
Db 694 ArgAlaGlyAla***AlaThrTrpTrpThrAlaProLeuAlaProSerArgIleGlyValAla 713
QY 1292 GGCAGCTGAGCGCGGCTACTTCCGCGACCTGTGTGTGACCGCGCGCGAGGCGCG 1351
Db 714 AlaAlaValArgProHisArgProProAlaArgProArgGlyLeuAlaProAlaHisAla 733
QY 1352 ATACGCGACCTTGCAGACACCTTACCGAGCGCGCGCGCATCATCTCGTACGCA 1411
Db 734 CysProProProThrArgProGlnSerArgAlaProAlaGlyAlaArgGlnGlnGlnGln 753
QY 1412 ACGGCGCGCGTGTGGCAAGAGAGAGCGGTTCACCGCGCGCGCGCGCGTGTGCTGC 1471
Db 754 ValAlaArgArgLeuGlyAlaGlyLeuArgSerProArgAlaAlaProArgArgArgGly 773
QY 1472 CACGCAAGCGCGCTGAGCGCGCGCGCGCGCTTACAAATCGCGGTACAGCGCGCGCT 1531

```

---

```

Db 774 ArgProCysProThrSerProGlnCysArgAlaAlaGlnPro-----GlyArgArg 790
QY 1532 GCGCGCGCGCGCCACCTTGGAGCAACCGCTACATGGCCCTCCCTCGCTCGCAATA 1591
Db 791 GlyGlyArgCysGlyProGlyThrAlaGlyGlyThrSerArgProSerGlyProCys 810
QY 1592 CGGCGCGCGC----- 1600
Db 811 ArgProArgAlaValThrThrAlaProPheLeuGlnProThrAspProAlaAlaProSer 830
QY 1601 -----CCGATATCGTGGCGCAAGAGTGAATGAGGCGCGCGCTGCGCG 1642
Db 831 SerArgSerArgSerArgProArgSerTrp-----ArgThrCysArg 844
QY 1643 CCGAGCGCGCAAG-----CCGGAATAATACCTTGCAGACCTGTGCGAGCGCAGC 1693
Db 845 CysSerValArgSerSerTrpProGlyGlyArgProCys***ThrArgProGlyProGly 864
QY 1694 GCATGCGCGCTCGA-----CCGTGTCAGAGCGCGAGCTGCGCGCAGA 1735
Db 865 Ala---ArgAlaArgValThrLeuProCysProAlaPro---TTPPTArg 879

```

RESULT 10

ID	ABB56523	standard; Protein; 1081 AA.
XX	ABB56523;	
AC	ABB56523;	
DT	05-MAR-2002 (first entry)	
DE	Human NMDA receptor subunit SEQ ID NO 44.	
XX		
KW	Human; N-methyl D-aspartate receptor; NMDAR; genetic screening;	
KW	glutamate receptor; drug; modulator; receptor.	
OS	Homo sapiens.	
XX		
PN	US6316611-B1.	
PD	13-NOV-2001.	
XX		
PF	29-SEP-1997; 97US-0940035.	
XX		
PR	20-APR-1994; 94US-0231193.	
PR	20-APR-1993; 93US-0052449.	
XX		
PA	(MERI) MERCK & CO INC.	
XX		
PI	Daggett LP, Lu C;	
XX		
DR	WPI: 2002-096550/13.	
DR	N-PSDB; ABA05910.	
XX		
PT	Novel nucleic acid encoding human N-methyl D-aspartate receptor 2A	
PT	subunit, useful for expressing the subunit and as probes in genetic	
PT	screening techniques, to identify patients having neuropathological	
XX	conditions	
PS	Example 3; Column 265-272; 206pp; English.	
XX		

The invention relates to nucleic acids (I, AI199167-AI199196, ABA05910) encoding human N-methyl D-aspartate receptor (NMDAR) subunits. The nucleic acids are useful for the production of NMDA receptor subunit proteins (ABB56503-ABB56530, ABA47961), as probes and to identify and isolate nucleic acids encoding related receptor subunits. Genetic screening can be carried out using the nucleotide sequences as probes to determine if any abnormalities exist with respect to any of the endogenous glutamate receptors. Also, patients having a family history of disease states related to glutamate receptor dysfunction can be screened to determine if they are also predisposed to such disease states. Expression vectors comprising (I), receptor subunits encoded by (I) and host cells comprising (I) are useful for producing selected NMDAR subunits and in specific combinations, as well as antibodies to the

CC receptor subunits substantially free of contamination from many other  
 CC receptor proteins whose presence can interfere with analysis of a single  
 CC NMDAR subtype. The availability of desired receptor subtypes allows the  
 CC observation of the effect of a drug substance on a particular receptor  
 CC subtype or combination of NMDAR subunits and to perform initial in vitro  
 CC screening of drugs or modulators in a test system that is specific for  
 CC humans, a human NMDAR subtype or the combination of NMDAR subunits.

XX Sequence: 1081 AA;

# Alignment Scores:

Pred. No.: 2,11e-18 Length: 1081  
 Score: 370.00 Matches: 189  
 Percent Similarity: 32.63% Conservative: 58  
 Best Local Similarity: 24.97% Mismatches: 234  
 Query Match: 11.22% Indels: 276  
 DB: 23 Gaps: 43

US-10-009-782-1 (1-1758) x ABB56523 (1-1081)

QY 32 CCAATGCTCC-----AATCGGAT-----CCGACG 55  
 Db 210 ProCysProAlaAlaGlyArgAlaThrThrProSerAlaAlaGlyThrTTPProThr 229  
 QY 56 CCTGACCTGCTGCTGCGCGCGGACCCATGACGCGACACCCCGG---GGC 112  
 Db 230 ProSerSerAlaValArgAspSerAlaSerThrSerArgSerTTPProGlyTTPSer 249  
 QY 113 GCGCGCGCG-----ACCTG-----GCGTGGCGCGC 139  
 Db 250 AsnSerProThrThrCysThrTTP\*\*\*ProThrAlaSerMetAlaSerGlyCysAlaAla 269  
 QY 140 ACCGACATCGCGCGCGCGCGATGTCGACGCGCGCGCGACACCCGGGTGACGT 199  
 Db 270 TyrGlyThrAla\*\*LeuGlyArgCysThrThrSerGlyGlnThr----- 284  
 QY 200 CGGAGCTGCTGCTGCGCGCGGCTCATCGACTGCG---ACACCCAGCAGACACTAC 256  
 Db 285 -----TTPProSerAlaProSerProSerMetAlaArgAlaProArgSer\*\*\*ThrSer 302  
 QY 257 TGTCTAGGCGCTCGCGACATGACGCCACCAAGATCTGCGAGCGCGCGCGACAGGTGACG 316  
 Db 303 Leu-----TyrProLeuThrAlaArgAlaSerVal\*\*\*TTP----- 314  
 QY 317 GCATTTGCGGACATCAGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 Db 315 -----TTPLeuAlaAlaMetAlaProSerProProArg 325  
 QY 374 ACCTGCTGACGAGAGCGGCTCTTACGCTTGAGCGGCTTCCGCGCTACCTGAGCGGT 433  
 Db 326 ProSerTTPSerHisThrLeu-----Gln 334  
 QY 434 TGGCG 493  
 Db 335 CysGly\*\*\*\*\*CysLeuSerCysAlaSerLeuTTPTrpPro-----Ser 349  
 QY 494 CCGCGCTCATCG 547  
 Db 350 ProSerSerCysSerSerThrSerAlaLeuSerAlaThrThrArgThrSerProAla 369  
 QY 548 -----TGGCGCGCGCGCG 559  
 Db 370 ArgSerProGlyAlaGlnLeuSerLeuSerAlaSerProCysGlyCysGlyArgTTP 389  
 QY 560 -----CCGAGAGACGATG--- 574  
 Db 390 SerSerThrThrGlnCysProSerArgThrArgGlyAlaProProAlaArgSerTTPhe 409  
 QY 575 -----CCAGCG 619  
 Db 410 TTPSerGlyProSerLeuLeuSerSer-----SerSerProAspThrArgProThrTTP 427  
 QY 620 CCG 643

Db 428 ProProSer\*\*\*SerLySerAsnThrSerThrLeuCysArgAlaSerValThrArgSer 447  
 QY 644 -----TCATCGAGGTG-----GCCGC 661  
 Db 448 PheSerGlyLeuLysIleSerThrHisLeuSerAlaSerAlaGlyCysProThrAlaAla 467  
 QY 662 CGCTGAGCG 721  
 Db 468 ArgSerGlyThrSerAlaValThrThrValThrCysThrProThrTTPSerSerThr 487  
 QY 722 TGGCGCGCTGAGG-----AACTTCCGATG 751  
 Db 488 SerAlaArgTTPArgThrArgSerProAlaSerArgTTPAlaLeuArgLeuSerLeuSer 507  
 QY 752 GCGCGAGCTGG-----ACGTGCGGTGATCTCC 784  
 Db 508 ProGlySerTTPMetProSerSerMetLeuLeuSerSerThrThrTTPGlnAlaArg 527  
 QY 785 ACACAAAGTCAATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 823  
 Db 528 ThrArgAlaAlaSerTTPSerProLeuGlyLeuAlaArgSerLeuLeuProLeuAlaThr 547  
 QY 824 GCGAGACGCTGCG 862  
 Db 548 AlaSerProCysArg-----ArgThrProThrGlySerGlyPro\*\*\*ThrTTPArgSer 565  
 QY 863 AGACGCT----- 871  
 Db 566 CysSerSerTTPGlyThrGlnArgAlaArgAsnTTPArgGlnCysGlySerGlnGlySer 585  
 QY 872 -----CGTGAAGCGT-----ATCCCTAGTGGCGCGCGCGCGCGCGCGCG 916  
 Db 586 AlaArgMetAlaArgThrArg\*\*\*\*AlaAlaSerTTPThrSerThrTTPGlnAla 605  
 QY 917 AGACCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 976  
 Db 606 SerSerThrCysCysTTP----- 611  
 QY 977 TGAGCG 1036  
 Db 612 -----TTPProTTPGlyTTPProCysTTPSerSerProGlySerThrTTP 626  
 QY 1037 TCCCGG---AGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1093  
 Db 627 SerThrGlySerCysAlaThrArgCysPro-----ThrHisProSerTTP 641  
 QY 1094 AGCGCATCTGCG 1150  
 Db 642 ThrSerCysTTPLeuSerAlaGlyAlaSerThrAlaAlaSerAlaGlyCysArgAlaSer 661  
 QY 1151 AGCG 1210  
 Db 662 ProAlaHis-----ArgGlyArgProAlaArgThrSer---ArgProAlaHisPro 677  
 QY 1211 ACCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1264  
 Db 678 ArgProAlaCysSerSerArgPheCysArgGln-----ProAlaThrTTP\*\*\*Pro 693  
 QY 1265 CCGCGCGCTGCG 1291  
 Db 694 ArgArgAla\*\*\*AlaThrProTTPThrAlaProLeuAlaProSerArgIleGlyValAla 713  
 QY 1292 GCGAGCTGACG 1351  
 Db 714 AlaAlaValArgProHisArgProProAlaArgProArgGlyLeuAlaProAlaHisAla 733  
 QY 1352 ATACCG 1411  
 Db 734 CysProProThrThrArgProGlnSerArgAlaProAlaGlyAlaYsArgGlnThrGly 753  
 QY 1412 ACG 1471

Db 754 ValAlaArgLeuGlyLeuArgSerProAlaAlaProAlaArgArgGly 773  
 QY 1472 CACGACGCGCGCTGAGCCGCGGCAAGCCCTTACATCCGCGTGAACGGGCGCGGT 1531  
 Db 774 ArgProCysProThrSerProGluCysArgAlaAlaGlyPro-----GlyArgArg 790  
 QY 1532 GCGCGCCCTCCCAACCCCTGAGCAACCGTACATGGCCCTCCCTCCGCTCCGCAATA 1591  
 Db 791 GtGtLArgGtGtGtProGtGtThraLgltGtThraArgProProSerGtGtGtGt 810  
 QY 1592 CCGCCCGCCAC-----CCGATATCGTGGGCAAGAGTATGGCGCGCGCTGCGCG 1600  
 Db 811 ArgProArgAlaValThraLThraLProPheLeuGluProThraAspProAlaAlaProSer 830  
 QY 1601 -----CCGATATCGTGGGCAAGAGTATGGCGCGCGCGCTGCGCG 1642  
 Db 831 SerArgSerSerArgSerProArgSerTrp-----ArgThrCysArg 844  
 QY 1643 CCGACGCGCAAG-----CCGGAATAATGACCTGCAAGACCTGTCGAGCGCGCG 1693  
 Db 845 CysSerValArgSerSerTrpProGtGtLArgProCys\*\*ThraArgProGtGtLArg 864  
 QY 1694 GCATCGCGGTCTCGA-----CCCTGTCGAAGCGCGAGCTGGCGCGAGA 1735  
 Db 865 Ala--ArgAlaArgValThraLeuProCysProAlaPro--TrpProArg 879

RESULT 11  
 ABG03533  
 ID ABG03533 standard; Protein; 900 AA.  
 AC ABG03533;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #3524.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.

XX WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PR 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSER-) HYSER INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YF;  
 XX  
 DR N-PSDB; AAS67720.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 20; SEQ ID No 33892; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. Ab000010-ABG03037 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SO Sequence 900 AA;

## Alignment Scores:

Pred. No.:	Length:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Indels:	Gaps:
2,5e-16	900	341.00	32.59%	28.16%	10.51%	272	41

US-10-009-782-1 (1-1758) x ABG03533 (1-900)

QY 1750 CGTACCTCAGGCGGCGATCTGCGCCAGCTGGCGCTTGAGAGGCTGACACCGCATGCCGC 1691  
 Db 9 ArgSerAlaArgSerValGtLArgArgPro-----GtThraLProAlaAlaArg 26  
 QY 1690 TGGCCTGAGACAGGCTTGACAGGTCATTTCCGGCCTTGCGCTCGCGCGCGCGCG 1631  
 Db 27 ProPro-----ArgProAlaAla 32  
 QY 1630 CCGCCATCACTCTCTCCCAAGATATCGGTGGCGCGCATATGCGAGCGAGGAGG 1571  
 Db 33 GlnArgGlnAlaLeuGtLysArgValGtLys-----Thrgly 46  
 QY 1570 CCATGATCGCGCTTGCGTCCAGG-----TTGGAGGCGGCGCGCGCGCG 1526  
 Db 47 ProGtLArgtLeuArgProGtGtLArgPheProGtLArgProArgGtLArg 66  
 QY 1525 CCGCGTTCAGCGCGGATGTAAG-----CGTGGCGCGCGCGCTCAGCGCG----- 1481  
 Db 67 AspAlaAspGlnArgAlaAlaArgGtLProValGtLArgGtLThraGlnHisGtLArg 86  
 QY 1480 -----CGTGC 1475  
 Db 87 AlaValProArgGtLProGlnAsnGluProAspGluThraLeuProGtGtLArg 106  
 QY 1474 GTGCGAGCA-----CGCGCGCGCATCTGCGCGTGAACGCGT-----GCTTGC 1427  
 Db 107 ProArgGtGtLArgtLeuArgGtLArgSerGtLArgGtLArgPro\*\*SerLeuThra 136  
 QY 1426 AGACCGCGCGCGCTTGATGATACAGGAGATGATGCGCGCGCGCTGAGGTGT 1367  
 Db 127 GlyProAla-----GTTGC 129  
 QY 1366 CGAAGTGGCGGTATGCGCACCGTGC-----GTTGCACACCA-----CCA 1322  
 Db 130 -----ProGtLArgGtLArgGtLArg\*\*SerProSerPro 142  
 QY 1321 GGTGCGAGATACCGCGCTGACAGTGCCTGCGCGCGCGCGCGCGAGCGCGG 1262  
 Db 143 GtLArgAlaSerSerLys--AlaGtLProTtLArgProGtLArgAlaSer 161  
 QY 1261 TCAGCGCGGTCACTTCATACCGCGCGCTCCAGG----- 1226  
 Db 162 LeuGlnArgAlaSerSerMetProAlaSerGlnValAspTtLArgGtLArgGtLArg 181  
 QY 1225 -----GGACAGCGCGCGAGTTCGCGCGCATAGT 1199  
 Db 182 ProArgCysAsnArgCysArgGtLArgLysProGtGtLArgGtLArgProArg 201

```

OY 1198 GCCCAGCACCAGCGGGAAGTGCCTC-----ACAGCGCGGATCGGGCGCT 1151
Db 202 LeuArgSerProGlyAlaLeuArgProGlyValGlyGlyLeuGlyLeuAlaLeuProAla 221
OY 1150 CGTGGTGGGAGAGCCCGTGGAGCCGATCATGTGCGGCCGCAACCCAGATGCGCTGCA 1091
Db 222 ArgThrAlaAla-----AlaAla 227
OY 1090 CGTGGGTGCTCCATCATGATGATGCGCGCGCGGCTGCA-----GCT 1043
Db 228 ProArgProArg-----GluArgTrpArgSerProGlyAlaProCysLeuGlyAla 244
OY 1042 CGGCGACACCGCTGCTACTTGATTTGCGCGCTGCGCGCATCTCATCAGATGCGCGCC 983
Db 245 Glu**Pro-----SerLeuProAlaAlaAlaGluProArg---GlyProGlyThrSer 261
OY 982 CGCTCAGTGGGAGGAGGCTGCACCGCATGATGATG----- 944
Db 262 -----GlyArgAla-----TrpProSerSerAlaAlaAlaArg, 272
OY 943 -----TGCCTCCGCGCAGCAGCAGCGGCTGCTTGAGCATGTTGAGCGCG----- 896
Db 273 GluArgPys---ArgAlaProGlyArgGlyProAlaAlaProThrGlyAlaArgArgPro 291
OY 895 CCAAGT-----AGGATACCGCGCTCCAGCGAGACGCTGCGC--- 860
Db 292 ProArgProGlyAlaGlyLeuAlaGlyArgProThrArgAlaAlaLeuProGlyArg 311
OY 859 -----GGCCCATGGCGG-----CCTGATCAGCGCGAGCGCTCGCGCAGC 818
Db 312 IleGlyArgArgSerProGlyArgAlaAlaProHisSerProAlaGlyLeuArgSerAla 331
OY 817 GGCAGAAATTGGCTGGCCCATGACTTGTGTGCGAGATCACACCG----- 770
Db 332 AlaArgArgCysProProProGlyProCysAlaProArgGlyGlnProGlyLeuProArg 351
OY 769 -----GCACGTCCAGCTCGC 755
Db 352 GlyGlyGlyAlaAlaAlaProLeuArgHisGlyGlyGlyAlaGlnLysThrArgProSerGln 371
OY 754 GGCAGATGGGAGAGTTTCCTCCAGCGCGCGCGAGATGCTCTGCTCGCGCATGT, 695
Db 372 GlyArgProGlyProProProProAlaProAlaGlyGlnArgGlnProGly 391
OY 694 GGGTGGCGTAGATGCGCGCATGCGCGCTCAGCGCGCGACACCTCGATGATCTCTGCG 635
Db 392 LysGluSerGlyGlnLysGlyValGlnProAlaArgGlyGluProGlnGlyProGlnArg 411
OY 634 TGGTGGCGGGCGCGCGCGGTAGAGCGCGCTCGAATGCGGATGCGCGCTGCG 575
Db 412 AlaMetArgGlyAlaAlaProAlaGlyArgArgPro----- 423
OY 574 CCATGGCTTCCTCGCGCAGTCCCGCATGCGCGCATTTCTCGTGGTGGCGCGCGCT 515
Db 424 -----ProAlaProAlaProGlyAlaAlaAlaGlnProGlyArgTrpGlyThrAla 440
OY 514 GCAAGTCCGCGCATGACCGCGCGCGCGAGCTGGAATGCGCCACCATACAGCGCGCTGA 455
Db 441 AlaGlyAlaAlaAlaGlyPro-----AlaAlaGlyArgProAlaProGlyArgLysProSer 458
OY 454 CGGCGCGCGCGCGCGCGCGCAACGCGCTAGTGGCGGAGCGCTCGAAGCGTAAG 395
Db 459 ArgAlaAlaGluGlyProAlaAlaAlaAlaProGlyGluArg-----GlyGln 473
OY 394 AGC-----CGCCTTCGTCGACGACGAGTCCAGGGGGG 365
Db 474 SerLeuProValProIleGlyValAlaGluGluArgGluGlyProArgGlyPro 493
OY 364 CGGCGCGGTGGCGCGCGCGCGCGCGCGCATGCGCGCAATTGGCGCGGACACCG 305
Db 494 ThrAlaAla---AlaGlyProAlaAlaAlaProGly----- 503
OY 304 TGGTGAAGCCCTCGCAGATCTTGGCGCTCATGTGCGCAGCGCTGAGCAGTAGTTGCTGT 245

```

```

Db 504 ----- 512
OY 244 ---CGTGGGTGGCGAGTCGATGAAAGCGCGCGCGACACCGCGCGACGTCGACCC 188
Db 513 AlaArgGly-----ProGluProAlaAlaProAlaArgGluPro 525
OY 187 GGGTGTGGCGCGCGCGCGCGACAGATGCGC-----CGA 155
Db 526 GlyArgArgAlaAlaGluLeuProArgThrArgGlyProProAlaProAlaProSerGluPro 545
OY 154 TGGCGCGCATGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104
Db 546 TrpProLeuProSerArgProProGlySerGlyLysArgGlyAlaSerAlaProProGly 565
OY 103 TGTTCCTCCGCTCGA-----TGAGCG 83
Db 566 GlyCysProSerArgSerGlyAlaSerProThrLeuGlnAspProGlyProHisSerGly 585
OY 82 TGGCGC-----CCGCGAGCAGCAGGT-----CGAAGGCTGGGAAATCGGATGGG 38
Db 586 CysProGluGluGlyPheProArgLeuAlaGlyGlyGluArgSerSerGlyProGluMetCyl 605
OY 37 ACA 35
Db 606 Thr 606

RESULT 12
ID AAB59827 standard; protein; 1592 AA.
AC AAB59827;
DT 04-APR-2001 (first entry)
DE Protein #4 encoded by Tuid/E gene.
DE XX
KM Toluene degradation; enzyme; waste degradation; Tuid; Tuid.
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azarcus toluylticus.
XX
PN MO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000MO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNITV OHIO.
XX
PI Coschignano FW;
XX
DR WPI; 2001-041080/05.
XX
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tuid (see AAF23629 and AAB59831), tuid (AAF23630 and AAB59832),
CC tuid (AAF23631 and AAB59833) and tuid (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, Tuid/E.

```

192 SerArgAspArgSerArgAlaSerArgSerArgGlySer-----ProLeuCysGlyAla 809

1128 CysA r g a r g h i s A n a r q a r q T y r g j v S e r a r a r q - D u o l -  
||||| . ||||| ||||| |||

[illegible]



[illegible]

OS-10-009-782-1 (1-1758) x AAB59826 (1-1615)

OY	14	TCGCGAAGAGAGAGATTTCCATTGCCAATCCGATCCAGC-----	CGT	58
Db	525	SeTleuAghAghAghAglyProSeTAlaArgSeTAlAlAlaIglyAenThHISLeuPro	11	54
OY	59	TCGACCTGCTGCTCCGGGGCGCA-----CCCGCATCGAGCGAGACACCCCGGGC	11	55
Db	549	AlaAenSeTleuAghAghAlaSeTArgProAlaSeTSeTAlaCysAlaIleProLysHIS	56	
OY	113	GGCGCGCGGACCTGGCGCTGGCGCGAGCAGCATCCGCGGCATCGGACGATCTGTGGAG	17	
Db	569	ThAArgPro-----AlAlYAlAlaenProLysProSeTAlAlaArgTTP	58	
OY	173	CGCGCGCGCACACCCGGGTGCGAGTGTCCGGCGCTGTGTGTGCGCGCGCT--TCATCG	22	
Db	584	AlaTTPArgThr-----SeTThSeTArgProTTPCysSeTArgThAenSeT	60	
OY	230	ACTGCACACCCACGAGACAGACACTACTGCTGTCAAGGCGTGGCAGCATGACGCCAAGNCT	28	
Db	602	SeTAlaThThThProLysIleProThnCysSeT-----ArgYs	61	
OY	290	CGCAGGCGTCACACAGGTGTCTACAGCGGCAATTGCCGATCAGACCTGGCGCGCTGGCGC	34	
Db	615	IleProAenCysProThnTTPProCysAghThThCysAglyAlaThThThArgYseT	63	
OY	350	ACGCCAACCGCG-----	36	
Db	635	ArgProThThArgThArgThArgSeTleuThnThIglySeTArgIAlaCysArgAlaSeT	65	
OY	362	CCGCGCCCCCGAGACCTGCTGAGACGAAGCGCGCTTACCGTTTCAGCGCTTCCGCACT	42	
Db	655	ValSeTProIleSeTlleArgIenThSeTAlAlaCysIleArgSeTAlAlA-----	67	
OY	422	ACGTGACGCGTTGCGGGCAGCGCGCGCGCGCTCAAGCGCGCGCGCTGATGTGGCCATT	48	
Db	673	--TTP-----ArgThArgProSeTleuPro-----	68	
OY	482	CAACGCTGGCGCGCGCGGTCATCCCGGACTTGCAGCGCGCGCCACGACGAGAGAAATCG	54	
Db	681	-----ValThThThAlaSeTcysArg-----ProThLysProSeT	69	
OY	542	CGGCCATGCGGAGACCTGGCGCGAGAGAACATGACGAGCGGCGCCATTCGATTTCAGCG	60	
Db	693	TTPLySThIglyCysTTPArgAlaSeTSeTSeTProLysSeTlleSeTPro--LysPro	71	
OY	602	GGCGCTTACCGCGCGCGCGCGCGCGCCGCGCCACACCGAAGATCANTCAGAGTGTGCCGC	66	
Db	712	ArgProThnCysAghProSeTProGly-----	72	
OY	662	CGCTGAGCGCATGCGCGCATACGCGACACCATGCGCGAGAGAGGAGACCATCG	72	
Db	721	-----ThAlaThArgThAlaSeTThTh-----	73	
OY	722	TGGCGCGCTGGAGAAACTTCGCAATCGCGCGAGCTGAGCTGCGGTGTGATCT	78	
Db	729	SeTProAgh-----SeT	73	
OY	782	CGCACACAGTGCATGGCCAGCCCAATT-----TCGCGCGCTGCGCGAGAGCG--	83	
Db	733	ThThThIglyAghArgTTPSeTSeTProAlaArgSeTleuIlyArgIAlaIlyArgAla	75	
OY	833	---TGGCGCTATGAGAGCGCGCATGCGCGCGCAGAGACGTCTCGCGAGCGGTATTCCT	88	
Db	753	GlyCys-----AlaArgSeTSeTArgIySThrSeTArgPro	76	
OY	890	ACGTGCGCGCTGCACCATGCTCAAGCAGAGACCGCGTGTGCTGCGCGGACGCACCATCA	94	
Db	765	IleArgSeTAlaArgProSeTcysSeTlys-----SeTProThr	77	
OY	950	TCACCTGTGCAAAGCCGTTCCCGCAACTGAGGCGCGACCTGGATGAAGTCGGGGCG	100	
Db	778	SeTAlaSeTAlaHePro-----Pro	78	

QY	1010	AGCGCGGCAAAATTCAGATGACAGCTGGTGGTCCCGAGCTGACGGCGCGGCGGCACTACT	1065
Db	785	SerProAlaArgAlaSerAlaThrArgCysAlaArgAsnSerLeuPro-----	800
QY	1070	TCATGATGAGCAACACCGGACGTGACGCGCATCTTGGCTTGGCCCGACATGATCGGCT	1129
Db	801	-----SerValThrArgSerSerAlaThrArgAlaAlaThrProArgArgIys	817
QY	1130	CCGAGCGCCTGGCCGACAGCAGAGGCCCGGACATCCGCGCTGTGGGACCTTCCCGCGGG	1189
Db	818	ThrProCysCysGlyAlaArgThrThrArgProProSerSerThrArgAsnSerSerArgAla	837
QY	1190	TCCTGGGCGACATATCGCGCGACCTGGGCGCTGTCCCGCTGGAGACGGCGGTATGAGA	1249
Db	838	ThrTrp-----MetArg-----TyrAsnSerSerArgTyrPasnValArgPhe-----	851
QY	1250	TGACCGCGCTGACCGCGCGCGCTTGGCGCTGGCCCGCGGCGCGGCGAGCTGCAGG-----	1303
Db	852	---ProSerMetAlaPro---AlaSerArgAlaProThrAlaIlysSerSerArgGlyArg	869
QY	1304	-----CCGGGTACTTGGCGCGACCTGTGGTGTGTGACACCGGCGCGAG	1345
Db	870	ThrlleCysSerSerSerProSerAlaAlaProThr-----ProArgAlaArg	885
QY	1346	TGGCGG-----ATACGCCACCTTGCAGACACCTACCGAGC-----GCG	1384
Db	886	ThrProAlaThrThrProThrProSerSerArgIleProSerGlySerAlaArgProSer	905
QY	1385	CCGCGCGGCAATCC-----ATTCCG-----TGTACG-----	1408
Db	906	ProProSerSerSerAlaIleProArgArgThrAlaArgArgArgCysAlaGlyPheSer	925
QY	1409	-----TCACGGCGCGCGCGTCTGGC-----AAGACAGCGCTTCA	1444
Db	926	SerAlaSerAlaThrAspSerSerAlaIleArgArgSerSerThrThrArgSerAlaArgSer	945
QY	1445	CCGGCCAGCATGCCGCGCGCGCTGCGACGACGAGCGCGCGTACGCCGCGCGCAACCTT	1504
Db	946	ArgArgAsnThrProSerSerSerAlaSerThrAlaThrAlaPro--ProThrArgIysPro-	964
QY	1505	TACAAATCCCGGCGGAACGGGGCGCGCTGCCCGCCCTCCCAACCTGACGCAAAACCGCT	1564
Db	965	--ThrThrGlySerThrCysCysAlaCysArgProAlaSerThrValAlaAlaArgArgL	984
QY	1565	ACATGGCCCTCCCTCCGCTCGCATATAGG---CCCAACCCGATATCGTGGCAAGAG	1621
Db	984	IysIys-ProValArgIysValAlaAlaGlnSerSerArgProSerCysTrp-----	1000
QY	1622	TGATGGCGCGCGCGCTGGCGCGCGGACCGGACCGGCAAAATGAGCC---TGCAGAGC	1678
Db	1001	-----LysSerArgSerMetThrAlaThrThrGlyArgThrProThrCysAsnSer	1017
QY	1679	TGTCGACGAGCCAGCGGACATCCGGGCTCGAGCC	1711
Db	1018	AlaArgArgProValIleSerArgArgSerPro	1028
RESULT 15			
ID	ABG03530	standard; Protein; 599 AA.	
XX	AC	ABG03530;	
XX	DT	13-FEB-2002 (first entry)	
XX	DE	Novel human diagnostic protein #3521.	
XX	KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	KW	food supplement; medical imaging; diagnostic; genetic disorder.	
XX	OS	Homo sapiens.	
XX	FN	MO200175067-A2.	



```
QY 1176 CACCTTCCGCG-----GGTGGGGGCACTATGCCGCGACCTGGCGCTGT 1223
Db 425 ArgArgSerAlaAlaHisGlyHisGlyAlaArgAlaGlySerProGlnProGlyAlaPro 444
QY 1224 CCGCGTGAAGACGCGGTATGAAAGATGACCGGCT-----GACGCGCGCGCTT 1274
Db 445 ProGlyHisLeuProGlyLeuProAlaArgGlnProLeuGlyLeuProAlaArgGlnArg 464
QY 1275 CCGCGTGGC-----CGGCGCGCGGCACTGCAAGCGCGGCTACTGCGCGACCT 1322
Db 465 CysPheGlyGlyLeuAlaGlnArgGlyArgAlaAlaArgHisCysLeuLeuSerArgPro 484
QY 1323 GGTGGTGT-----CGACCGCGGCAAGGTGCGCGATACGCG 1358
Db 485 SerAlaLeuAlaLeuArgHisSerSerTyrArgGlnProGlyMetGlyGly----- 501
QY 1359 CACCTTGAACACCTTACCGAGCGCGCGCGCATTCCTCGTACGTACGTCACGCGCGC 1418
Db 502 -----TTPArgSerProGlnAlaLeuGlyGlyTyrGlyGlySerGln 516
QY 1419 GCCGCTGTGGCAAGACGCGCTTACCGGCGCAAGCATGCGCGCGCTGCTGCAAGCAC 1478
Db 517 AlaGlySerAlaAlaGlyLeuSer-----GlyAlaAlaSerGln 528
QY 1479 GCGCGCTGAGCGCGCGCGCACCCCTTACAATCCGCGCTGAACGCGGCGCGCTGCGC 1535
Db 529 GlyArgArgAla-----ArgHisLeuArgGlyLysAlaAlaProAla 541
QY 1536 -----CCCCCTCCACCCCTGAGCGCAACCGCTACATGCGCCCTCCCTCCGC 1583
Db 542 TrpAsnProAlaProProSerPro---ProProProAlaLeuGlyLeuProLeuArg 560
QY 1584 TCGCAATACGCGCGCGCATATCGTGGCAAGAGTGAATGGCGCGCGCTGCGCGC 1643
Db 561 ThrGlnArgGlnAlaLeuArg-----LysProArgArgGlnGlnAlaArg 575
QY 1644 CGAGCGCAAGCGCGGAAATGACCTGCAAGACCTGCGAGGCGCAAGCATCGCGGT 1703
Db 576 Arg-----ProArgProArgProLeuArgPro----- 584
QY 1704 CTCGACCTGTCCAAAGCGCGAGCTGGGCGCAAGTCGC 1739
Db 585 ---GlyGlyAlaAnGlySerProGlyProProArg 595
```

Search completed: May 11, 2003, 12:09:19  
Job time : 97 secs

GenCore version 5.1.4.p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 01:22:49 ; Search time 28310 seconds

(without alignments)  
1005.711 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 1758  
Sequence: 1 gaattcacttgatcgagcga.....ccctgagctacgagaagctt 1758

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

EST:  
1: em\_estha:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_Other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Match	Length	DB	ID	Description
1	83	4.7	1516	12	BG809984	BG809984 mgct002xd
2	82.8	4.7	1798	17	AG171124	AG171124 Pan trogl
3	82	4.7	1798	17	AG171124	AG171124 Pan trogl
4	78.4	4.5	1542	17	AG032943	AG032943 Pan trogl
5	76	4.3	1189	17	AG030608	AG030608 Pan trogl
6	74	4.2	1116	12	BG810038	BG810038 mgct002x1

7	72.4	4.1	925	17	CNS0091P	AL053013 Drosoph11
8	72.4	4.1	1299	17	AG039481	AG039481 Pan trogl
9	71.8	4.1	932	17	CNS00720	AL066742 Drosoph11
10	71.8	4.1	1341	17	AG030611	AG030611 Pan trogl
11	71.6	4.1	925	17	CNS0091P	AL053013 Drosoph11
12	71.2	4.1	1152	17	AG076818	AG076818 Pan trogl
13	71.2	4.1	1538	17	AG030607	AG030607 Pan trogl
14	71	4.0	1129	17	AG127788	AG127788 Pan trogl
15	71	4.0	1281	12	BG852363	BG852363 10240340
16	71	4.0	1288	14	B0678719	B0678719 AGENCOURT
17	70.2	4.0	935	17	CNS006XK	AL066051 Drosoph11
18	70.2	4.0	1131	17	AG042920	AG042920 Pan trogl
19	69.6	4.0	1625	17	AG043477	AG043477 Pan trogl
20	69.2	3.9	1360	13	BM463178	BM463178 AGENCOURT
21	69	3.9	1160	14	BM926731	BM926731 AGENCOURT
22	68.6	3.9	935	17	CNS006XK	AL066051 Drosoph11
23	68.6	3.9	1364	14	BM810045	BM810045 AGENCOURT
24	68.4	3.9	932	17	CNS00720	AL066742 Drosoph11
25	68.4	3.9	982	14	B0687717	B0687717 AGENCOURT
26	68.4	3.9	1626	10	AW731151	AW731151 GA_Ea001
27	68.2	3.9	1387	10	AW731151	AW731151 GA_Ea001
28	67.8	3.9	936	12	BG852371	BG852371 10240340
29	67.8	3.9	1232	17	AG072425	AG072425 Pan trogl
30	67.8	3.9	1328	17	AG043615	AG043615 Pan trogl
31	67.6	3.8	1088	17	AG075424	AG075424 Pan trogl
32	67.6	3.8	1299	17	AG039481	AG039481 Pan trogl
33	67.4	3.8	803	17	AG162326	AG162326 Pan trogl
34	67.4	3.8	1189	17	AG030608	AG030608 Pan trogl
35	67.4	3.8	1278	17	AG060116	AG060116 Pan trogl
36	67.4	3.8	1651	12	BG809816	BG809816 mgct001xk
37	67.2	3.8	644	12	BG807300	BG807300 WHE493_G
38	67.2	3.8	1248	17	AG049208	AG049208 Pan trogl
39	66.6	3.8	1040	12	BG786331	BG786331 SEAMC006
40	66.4	3.8	1366	17	AG065058	AG065058 Pan trogl
41	66.2	3.8	1086	17	AG077581	AG077581 Pan trogl
42	66.2	3.8	1101	17	AG039543	AG039543 Pan trogl
43	66	3.8	1164	17	AG128373	AG128373 Pan trogl
44	65.8	3.7	1339	17	BM458211	BM458211 AGENCOURT
45	65.8	3.7	1452	17	AG032979	AG032979 Pan trogl

## ALIGNMENTS

RESULT 1  
Bg809984  
LOCUS mgct002xd1f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe  
DEFINITION Bg809984 1516 bp mRNA linear EST 22-MAY-2001  
ACCESSION Bg809984  
VERSION Bg809984.1 GI:14180964  
KEYWORDS EST  
SOURCE Magnaporthe grisea  
ORGANISM Magnaporthe grisea  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetes; Incertae sedis; Magnaporthaceae; Magnaporthe.

REFERENCE 1 (bases 1 to 1516)  
AUTHORS Choi, W. and Dean, R. A.  
TITLE Construction and sequence analysis of an appressorium stage cDNA  
JOURNAL library in the rice blast fungus, Magnaporthe grisea  
COMMENT Unpublished (2001)

## JOURNAL

## FEATURES

source  
Email: ralph.dean@ncsu.edu  
Seq primer: T3 primer (AATTAACCTCACTAAAGG).  
Location/Qualifiers  
1. 1516  
/organism="Magnaporthe grisea"  
/strain="70-15"

Query Match	4.78;	Score 83;	DB 12;	Length 1516;
Best Local Similarity	40.78;	Pred. No. 0.00033;		
Matches 530; Conservative	0;	Mismatches 765;		

QY	948	CATCACCTGTGGTGCAGGACCCCTTCCCGAAGTGAAGCGGCGCAGCACTGTGATGAATTCGCGGC	1007
Dp	989	CCCCCGCGCGCGCGGCCGCCCGCCGCMGCGCCCGGGGAGACCCTCCGACGCGCGCGCGCGGCC	1048
QY	1008	CGAGCGCGGCAAAATCCAAGTACGACGATGTGTGCGCCGAGTGACGCGCGCGCGCGCGCATCTA	1067
Dp	1049	GCG	1108
QY	1068	CTTCAATGATGAGACAACCCGACGTGACAGCGCATCTGTGCGCGCGCGCGCGCGCGCGCGCGCG	1127
Dp	1109	GNGCG	1168
QY	1128	CTCCGAGCG	1187
Dp	1169	GCG	1228
QY	1188	GGTGCTGGGGGACATATGACG	1247
Dp	1229	ACNCGACG	1288
QY	1248	GATACCG	1307
Dp	1289	GCG	1348
QY	1308	GTAATTGCG	1367
Dp	1349	CCGCGACG	1408
QY	1368	ACAACCGACG	1408
Dp	1409	CCG	1449
RESULT 2	AG171124		
LOCUS	AG171124		
DEFINITION	Pan troglodytes DNA, clone: RP43-04F09.TJ, genomic survey		
ACCESSION	AG171124		
VERSION	AG171124.1 GI:16700802		
KEYWORDS	GSS.		
SOURCE	Pan troglodytes male lymphocytes DNA, clone_1lb:RP43 Chimpanzee Male BAC library clone:RP43-04F09.TJ.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
REFERENCE	1 Fujiiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tochik, Y., Watanabe, H. and Sakaki, Y. BAC end sequences of Library RPCI-43 Unpublished 2 (bases 1 to 1798) Fujiiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tochik, Y., Watanabe, H. and Sakaki, Y. Direct Submission Submitted (02-AUG-2001) Asao Fujiiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Shinhato-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpbes@gsc.riken.go.jp, URL: http://hnp.gsc.riken.go.jp/ Tel: 81-45-503-9111, Fax: 81-45-503-9170) Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated after the Rad process and may have higher chance of clone tracking errors. PRIMERS		
TITLE	JOURNAL		
REFERENCE	AUTHORS		
COMMENT			
FEATURES	source		
LIBRARY	Vector : pBACE3.6 R.site 1 : EcoRI R.site 2 : EcoRI Location/Qualifiers 1..1798		



[illegible]

RESULT 3  
AG171124/c

LOCUS AG171124 1798 bp DNA linear GSS 09-JAN-2002

DEFINITION Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey sequence.

ACCESSION AG171124

VERSION AG171124

KEYWORDS GI:16700802

SOURCE GSS

ORGANISM Pan troglodytes male lymphocytes DNA, clone\_1lb:RPCI-43 Chimpanzee Male BAC Library clones:RP43-040F09.TJ.  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Carnivora; Vertebrata; Euteleostomi;  
Mammalia; Euarchonta; Primates; Catarrhini; Hominoidea; Pan.

REFERENCE 1

AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Tookki,Y., Watanabe,H. and Sakaki,Y.

TITLE BAC end sequences of library RPCI-43 unpublished  
2 (bases 1 to 1798)

JOURNAL

REFERENCE

AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Tookki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

Submitted (12/30/2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-1-1, Hirosaki, Aomori-shi, Aomori, 030-0192, Japan  
E-mail: fujiyama@riken.go.jp, yorokama@riken.go.jp  
Tel: 81-45-503-9111, Fax: 81-45-503-9170  
Cloned and derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the RbD process and may have higher chances of clone tracking errors.  
primers

Vector : pBACE3.6

R.Site 1 : ECORI  
R.Site 2 : ECORI

1. .1798

86/T: T

```

/db_xref="taxon:9598"

```

```

/sex="male"
/sex="male"

```

/Cell\_Type="Lymphocytes"

778 c	655 g	0 t	365 others
-------	-------	-----	------------

4.78; Score 82; DB 17; Length 1708.

Frequency: 0.0003; Matches: 805; 1-3-1-1

-----

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

226

1384

ACCCACGACCACTACCTGCTCAGGCGTCGGCATGACGCCAAGA 286

1.324

GTCACCA CGGTG GTCACGGGCAATTGCGGCATCAGCCTGGCGCCGCTGG 346

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

.....

**F0ZT UNDCOCCCCCCCCCCCCCCCCCC**

466

**CCTGCGCCGCNGC** 1144


CAATTCACGCTGCGCGCCGGTCA TGCCGACTTGCAGCGCGCGCCA 526

5'NGGCGGCGCCGCGGNCGCGCGCGGGCGGCGGNCGCGGNGNGNCGCCG 1084

AATCGCGGCCCATGCGGGGACCTGGCCGAGGAAGCATTGCGCAAGCGG--CGCG E01

[illegible]

•



**\*06**

-----CAGGCGCCACCACTGGCGGA /04

905

ACATCTGGCCGCTGGAGGAACCTTCGCGATCGGCCGCGAGCTGGA 764

[illegible]



RESULT 6  
BG810038/c  
LOCUS BG810038 1116 bp MRNA linear EST 22-MAY-2001  
DEFINITION mgcct002x101f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe  
ACCESSION grisea cDNA clone mgcct002x101f 5', mRNA sequence.  
VERSION BG810038  
KEYWORDS BG810038.1 GI:14181018  
EST.  
SOURCE Magnaporthe grisea.  
ORGANISM Magnaporthe grisea  
REFERENCE 1 (bases 1 to 1116)  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetes; Incertae sedis; Magnaporthaceae; Magnaporthe.

[illegible]



ACCESSION AL066742  
VERSION AL066742.1  
KEYWORDS GSS.  
SOURCE GI:4945205  
ORGANISM *Drosophila melanogaster*.  
*Drosophila melanogaster*.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;

RESULT 10		
AG030611/c		
LOCUS	AG030611	
DEFINITION	Pan troglodytes DNA, clone: pIB-00320.F, genomic survey sequence.	
	1341 bp	linear
		GSS 01-NOV-2001



[illegible][illegible]

```

Db 805 GGASASHSSSSACBSSSSSCSASCMWSSSSSSASRSRSGAGAGSASSSSSSSS 746
Oy 1314 GCGGACCTGCTGTGTGACCGGCGGACGCTGGCCCATACCGCCACTTCGACACCC 1373
Db 745 ASAGSVSSASASSSSSSSSSVSCSVASSMSSSSSSASASSSSSSSSASCAACSC 686
Oy 1374 TACCGAGCGCGCGCGGACATCCATCCGTACGTCACGCGCGCGGCTGCGCAGA 1433
Db 685 TSMGCSCTASMSASMSASSSSSSSSSSSSASASSSSSSSSSSSSSSSSSSSS 626
Oy 1434 GCAGCGCTTACACCGCGCGGACATCCGCGCGCGGCTGCGCAGCAGCGCGCGCTGAGCCG 1493
Db 635 SMSGSGGSGSVASAGSMSSSVSSSGGSGGSGGSGGSGGSGGSGGSGGSGGSG 566
Oy 1494 GCGGACCGCTTACATCCGCGGCTGGAACGCGGCGG 1527
Db 565 SGCMCRCSGSSAAAASCAVAAACGCMGCKSG 532

RESULT 12
AG076818. 1152 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-071C05.R, genomic survey sequence.
ACCESSION AG076818
VERSION AG076818.1 GI:16628620
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
ORGANISM BAC library clone:PTB-071C05.R.
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
TITLE Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1152)
DIRECT SUBMISSION
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
Location/Qualifiers
source 1. 1152
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-071C05.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 34 a 488 c 514 g 23 t 93 others
ORIGIN
Query Match 4.1%, Score 71.2; DB 17; Length 1152;
Best Local Similarity 45.3%; Pred. No. 0.028;
Matches 308; Conservative 0; Mismatches 368; Indels 4; Gaps 3;
Oy 105 CCGGCGGCGCGCGCGCGCGCTGGCGGCGGAGCGGACGATCCGCGCGGCGGAGACT 164
Db 389 CCGGCGGCGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 448

COMMENT
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.

Oy 165 GTGCGAGCGCGCGCGCGGACACCGCGGTGACGTGTGCGGCGCTGTGTCGCGCGCGCTT 224
Db 449 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 508
Oy 225 CATGACATCGCACACCGACGACGACGACGACGACGACGACGACGACGACGACGAC 284
Db 509 GGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 566
Oy 285 GATCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 344
Db 567 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 626
Oy 345 GCGCGACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 404
Db 627 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 666
Oy 405 CGAGCGCTTGGCGGATCTACCTGAGCGGCTTGGG -GCCAGCGCGCGCGCGCGCGCG 463
Db 687 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 746
Oy 464 CCGTATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
Db 747 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806
Oy 524 CCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 583
Db 807 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 866
Oy 584 CCATCGGCAATTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 643
Db 867 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 925
Oy 644 TCATGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 703
Db 926 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 985
Oy 704 ACGAAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 763
Db 986 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1045
Oy 764 ACGTCCGCGTGTGATCTCG 783
Db 1046 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1065

RESULT 13
AG030607 1538 bp DNA linear GSS 01-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-003A18.F, genomic survey sequence.
ACCESSION AG030607
VERSION AG030607.1 GI:16557480
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
ORGANISM BAC library clone:PTB-003A18.F.
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
TITLE Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1538)
DIRECT SUBMISSION
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.

Oy 165 GTGCGAGCGCGCGCGCGGACACCGCGGTGACGTGTGCGGCGCTGTGTCGCGCGCGCTT 224
Db 449 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 508
Oy 225 CATGACATCGCACACCGACGACGACGACGACGACGACGACGACGACGACGACGAC 284
Db 509 GGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 566
Oy 285 GATCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 344
Db 567 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 626
Oy 345 GCGCGACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 404
Db 627 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 666
Oy 405 CGAGCGCTTGGCGGATCTACCTGAGCGGCTTGGG -GCCAGCGCGCGCGCGCGCGCG 463
Db 687 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 746
Oy 464 CCGTATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
Db 747 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806
Oy 524 CCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 583
Db 807 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 866
Oy 584 CCATCGGCAATTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 643
Db 867 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 925
Oy 644 TCATGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 703
Db 926 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 985
Oy 704 ACGAAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 763
Db 986 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1045
Oy 764 ACGTCCGCGTGTGATCTCG 783
Db 1046 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1065

```



RESULT 15	LOCUS	DEFINITION	EST	EST 29-MAY-2001
BG852363	10240334A03.y2	1281 bp	linear	
	Chlamydomonas reinhardtii CC-1690, normalized, lambda Zap II			
ACCESSION	BG852363			
VERSION	BG852363.1			
KEYWORDS	GI:14233547			
SOURCE	EST.			
	Chlamydomonas reinhardtii			

BASE COUNT 30 a  
ORIGIN 697 c 456 g 15 t 83 others

/\*strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, normalized, lambda Zap  
/note="Vector: pBluescript II SK-; Site.1: EcoRI; Site.2:  
XhoI; This library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP (acetate-containing) medium in the  
light, TAP medium in the dark, HS (minimal) medium in the  
ambient levels of CO2 and HS medium bubbled with 5% CO2.  
POLYA mRNA was purified from each sample, pooled and cDNA  
synthesized. The cDNA was directionally cloned into lambda  
Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
Zap clones by superinfection with Exsist (Stratagene)  
phase. The library was normalized using method 4 described  
in Bonaldo et al (1996) Genome Research 6: 791-806."

Query Match	4.08;	Score 71;	DB 12;	Length 1281;
Best Local Similarity	42.88;	Pred. No. 0.031;		
Matches 515;	Conservative	0;	Mismatches 675;	Indels 14;
			Gaps	7;
QY	336 GGC	GGC	GGC	GGC
Db	70 GGC	GGC	GGC	GGC
QY	396 TTA	CGT	TTT	GAG
Db	130 CCG	CGC	CCC	CCC
QY	456 CA	AGC	CGC	CGT
Db	190 GCG	GGC	GGC	GGC
QY	516 GCG	GGC	GGC	GGC
Db	250 CAA	AGC	CGC	CGT
QY	576 GAG	CGC	CGC	CGT
Db	310 CAG	CGC	CGC	CGT
QY	636 GCA	AGC	CGC	CGT
Db	370 GGG	CGC	GGC	GGC
QY	696 CAT	GCG	CGC	CGT
Db	429 CCG	GGC	GGC	GGC
QY	756 CGA	CGT	GCG	CGT



GenCore version 5.1.4.p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:03:15 ; Search time 753 Seconds  
(without alignments)  
11044.401 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 1758  
Sequence: 1 gaattccactgacgcgcgga.....ccctgacgcgcgaagctt 1758

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 6135869 seqs, 2365311053 residues.

Total number of hits satisfying chosen parameters: 12271738

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_New.\*  
1: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*  
7: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*  
8: /cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq:\*  
9: /cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq:\*  
10: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq:\*  
11: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	483	27.5	US-10-282-122A-14027	Sequence 14027, A
2	425.4	24.2	US-10-282-122A-24070	Sequence 24070, A
3	161.2	9.2	US-10-156-761-3269	Sequence 3269, Ap
4	118.4	6.7	US-10-156-761-3783	Sequence 3783, Ap
5	116	6.6	PCT-US02-37547-12	Sequence 56, Appl
6	116	6.6	PCT-US02-37547-12	Sequence 56, Appl
7	95.4	5.4	PCT-US02-37547-12	Sequence 56, Appl
8	91	5.2	US-10-156-761-6189	Sequence 6189, Ap
9	90.2	5.1	PCT-US02-37547-12	Sequence 12, Appl
10	88.4	5.0	PCT-US02-37547-12	Sequence 12, Appl
11	85.8	4.9	PCT-US02-37547-12	Sequence 12, Appl
12	82.6	4.7	US-10-329-079-12	Sequence 12, Appl
13	82.6	4.7	US-10-329-079-12	Sequence 12, Appl
14	82.4	4.7	US-10-156-761-6160	Sequence 6160, Ap
15	82	4.7	US-10-156-761-6160	Sequence 6160, Ap
16	81.8	4.7	US-10-398-605-7	Sequence 1, Appl
17	81.8	4.7	US-10-398-605-7	Sequence 1, Appl
18	81.8	4.7	US-10-398-605-7	Sequence 1, Appl
19	81.8	4.7	US-10-398-605-7	Sequence 1, Appl
20	80.8	4.6	US-10-156-761-4343	Sequence 4343, Ap
21	80.6	4.6	US-10-156-761-1922	Sequence 1922, Ap
22	80.4	4.6	US-10-282-122A-25795	Sequence 25795, A

23	80.2	4.6	11817	8	US-10-156-761-2884	Sequence 2884, Ap
24	80.2	4.6	125746	8	US-10-156-761-15102	Sequence 15102, A
25	79.2	4.5	1458	8	US-10-419-128-1112	Sequence 1112, Ap
26	79.2	4.5	1458	8	US-10-366-683-1112	Sequence 1112, Ap
27	79.2	4.5	1461	8	US-10-419-128-972	Sequence 972, App
28	79.2	4.5	1461	9	US-10-366-683-972	Sequence 972, App
29	79.2	4.5	4563	8	US-10-419-128-930	Sequence 930, App
30	79.2	4.5	4563	9	US-10-366-683-930	Sequence 930, App
31	79	4.5	2280	8	US-10-369-493-42316	Sequence 42316, A
32	78.6	4.5	1404	8	US-10-156-761-3773	Sequence 3773, Ap
33	78.2	4.4	12249	9	US-10-267-255-74	Sequence 96, Appl
34	78.2	4.4	18331	9	US-10-267-255-96	Sequence 96, Appl
35	78.2	4.4	18876	9	US-10-329-079-42	Sequence 42, Appl
36	78.2	4.4	61944	9	US-10-329-079-34	Sequence 34, Appl
37	77.8	4.4	3759	8	US-10-156-761-2268	Sequence 2268, Ap
38	77.8	4.4	5304	8	US-10-156-761-3184	Sequence 3184, Ap
39	77.8	4.4	9025608	8	US-10-156-761-1	Sequence 1, Appl
40	77.4	4.4	4851	9	US-10-282-122A-25489	Sequence 25489, A
41	76.2	4.3	5970	9	US-10-201-365-11	Sequence 11, Appl
42	76.2	4.3	5970	9	US-10-160-5398-21	Sequence 21, Appl
43	75.8	4.3	9975	1	PCT-US02-37547-4	Sequence 4, Appl
44	75.8	4.3	82746	1	PCT-US02-37547-56	Sequence 56, Appl
45	75.6	4.3	1719	8	US-10-156-761-3854	Sequence 3854, Ap

## ALIGNMENTS

RESULT 1  
US-10-282-122A-14027  
Sequence 14027, Application US/10282122A  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Travick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA 034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 14027  
LENGTH: 1425



TYPE: DNA  
ORGANISM: Burkholderia mallei  
US-10-282-122A-14027

Query Match 27.5% Score 483; DB 9; Length 1425;  
Best Local Similarity 60.3%; Pred. No. 2,1e-66;  
Matches 866; Conservative 0; Mismatches 520; Indels 51; Gaps 2;

40 CAATCGATTTCCAGCCCTTCCACCTGCTGCTCGGGGGGAGACCCCTCATCGACGACG 99  
Db CACTCCGATCCCGAAGTGGCCGATAGCTGATGCTGGGGCGCAGATTGATACAGCGACG 63  
QY 100 AACACCCCGGGGGGCGGCGCCGACCTGGGGCGGCTGGGGAGCCGATGCGCCATCGGC 159  
Db GCGGCGCCGAGGCTGAGAGGGGACCTCCGATACCGGGGAGCCGATGCTGCGGATCGGC 123  
QY 160 GATCTGTGAGAGCCCGCGCGCACACCCGGGTGACGATGTGGGGCTGTGGTGGCGGCC 219  
Db 124 AATGTACGAACTGGCTTGGCGGAGAGGTGATGAGCGGAACGGCCGGCTGGCGGCC 183  
QY 220 GGCCTTCATGACTCGACACCCAGAGACAACCTGCTCAGGCGCTCGGACATGACG 279  
Db 184 GGCCTTCATGACTCGACACCCAGAGACAACCTGCTCAGGCGCTCGGACATGACG 243  
QY 280 CCCAAGATCTGCGAGGGGCTGACCAACGAGTGTGACAGGGCAATGGGATGAGCTGGCG 339  
Db 244 CCGAAGATCTGCGAGGGGCTGACCAACGAGTGTGACAGGGCAATGGGATGAGCTGGCG 303  
QY 340 CCGGCGCGCAAGGCGCAACCCCGCCCGCTGAGACTGCTGAGCAAGGCGGCTCTTAC 399  
Db 304 CCGGCGCGCTC-----AAGGCGGATCGGCC 329  
QY 400 CGTTTCAGAGCGCTTCGCCACTGACGAGCGCTTGGGGGCAACCGCGGGCGCTAC 459  
Db 330 CGATTCGA-----TGACCTGCTCGCGGGGGAAGCGCGCTAC 372  
QY 460 GCGGCGCTGATGCTGGGCGCATTAACGCTCGGCGCGCGGCTCATGCGGACTTGCACGC 519  
Db 373 GTCGCGGCGCTGCTGGCGCACACCGCGCTCGCGAGCAACAGATGAGCGCGCTGACGC 432  
QY 520 GCCGCCACGACGAGAAATCGGGGCGGAGCTGGCGGAGAGAGGAGCGGCGAGC 579  
Db 433 GCGGCGACGCGCGGCAATCGGGGCGGAGTGGCGGCGAGTCTGAGAGGCGCTGGCGAC 492  
QY 580 GGGGCGCATCGCATTTTCGACCGCGGCTTTCACCCCGCGCGCGCGCAACCGAA 639  
Db 493 GGGGCGCTCGCTTTCACCGCGCTTCGCTGACGGCTTCGCGCGCGCGCGCGAG 552  
QY 640 GAGATCATGAGGTGGCGCGGCTGAGCGGCGCATGGCGCATTAACCGCAACCGCATG 699  
Db 553 GAGGTATGCGCTCGCGGAGCGGCTGCGGAGAGCGCGGCGCGTCTACAGAGCGCATG 612  
QY 700 CCGGAGAAAGGAGACATCGTGGCGGCTGGAGAAACCTTCGCGATCGCGCGGAG 759  
Db 613 CCGACCGAGTTCGATCGATTCCTCGAGGAGTGAAGTGGCGGCGCTACCGCGCTGGCGCAT 672  
QY 760 CTGAGAGTCCGCTGTGATTCGCAACAAAGTATGAGGCGGCGCAATTCGAGCGC 819  
Db 673 GCAAGGCTCGCTGTGATTCGCAATCGAGTGGCGGCGCGCTGCAATGCGGCGC 732  
QY 820 TCGCGGAGACGCTGCGGTGATCGAGGCGCGGCGGCGGCGGCGGCGGCGGCGG 879  
Db 733 AGCGGAGAGTCTCGCGCTCGTCAAGCGCGGCGGCGGCGGCGGCGGCGGCGG 792  
QY 880 GCGTATCCCTTACGCGCGGCTGACACATGCTCAAGAGAGCGGCGGCTGCTGGCGGA 939  
Db 793 TCGTATCCCTTACGCGCGGCTGACACATGCTCAAGAGAGCGGCGGCGGCGGCGG 852  
QY 940 CCGACATCATCACTGCGGAGAGCGCTTCCCGAATGAGCGGCGGCGGCGGCGGCGG 999  
Db 853 GACATCAGATTCAGTGTTCGATCCGATCCGAAAGGCGGCGGCGGCGGCGGCGG 912  
QY 1000 GTGCGGCGCGGCGGCGGCAATTCAGATGAGTGTGCGCGGCGGCGGCGGCGGCGG 1059

Db 913 ATGCGCGCGGATGGCGGCTGTGCTCGAGAGAGCGGCGCGCGGCTCGACCGCGGCG 972  
QY 1060 GCCATCTACTTCATGATGAGAGACCGGAGCTGAGCGGCGGCGGCGGCGGCGGCGG 1119  
Db 973 GCGGCTGATCAAAATCTCGAGAGAGAGAGCTGCGCGGATGCTGTGCGACCGCGGCG 1032  
QY 1120 ATGATCGGCTCGAGAGGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1179  
Db 1033 ATGATCGGCTCGAGAGGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1092  
QY 1180 TTCCGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1239  
Db 1093 TTTCGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1152  
QY 1240 GTATGAAATGAGAGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1299  
Db 1153 GTGCGGAAATGAGAGCTGCGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1212  
QY 1300 CAGGCGGCTACTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1359  
Db 1213 CCGGCTCGCTATCACGCGGATCTGCTGCTGATGCGGCGGCGGCGGCGGCGGCGG 1272  
QY 1360 ACCTTCGAAACCCCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1419  
Db 1273 ACCTTCGAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1332  
QY 1420 CCGGCTGCGGAGAGAGCGGCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1476  
Db 1333 CTGTCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1389

## RESULT 2

US-10-282-122A-24070

Sequence 24070, Application US/10282122A

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: EITRA.034A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614  
 SOFTWARE: Patent version 3.1  
 SEQ ID NO: 24070  
 LENGTH: 1437  
 TYPE: DNA  
 ORGANISM: Klebsiella pneumoniae  
 US-10-282-24070

Query Match 24.2% Score 425.4; DB 9; Length 1437;  
 Best Local Similarity 57.4%; Pred. No. 1.8e-57;  
 Matches 787; Conservative 0; Mismatches 581; Indels 3; Gaps 1;

59 TCGACTGCTGCTGCGGGGCGACCTCATGACGCGGCAACACCCGGGGGGCG 118  
 8 TTGACTGCTTTTAAAAACGTCAAGGTGATGACGAGGAGGGGGCGCCAAATACCTG 67  
 119 CCGACTGGGGGGGGGGGGGGGCGCATGCGCCGCTGCGGATCTGTGAGCGCGCG 178  
 68 CTGATGTGGCGGTGAAGGCGCATGATTAATGCTATCGCCCGCGCTGAGACCTTGGCG 127  
 179 CGCAGACCCGGGTGACGTGTGCGGCTGTGTGCGCGCCCGCTTCACTGACGCGACA 238  
 128 CGGAGCAGGTGATGATGAGCCAGGGGGGGGTACTGGCGCGCTTATTTATGTGCTA 187  
 239 CCCAGCAGCAACTACTGCTCAGGCTGCGGACATGACGCCCAAGATCTGCGAGGCG 298  
 188 CCCATGATGATTAACAGTATCGCATGCGGAGTACTTCCGAGCTCAGCAGGCGG 247  
 299 TCACGACGCTGTGACGCGGCAATGCGCATGACCTGCGGGCGCGCTGGC---GACGCCA 355  
 248 TGACGACGGGTGATGTGTGGAACCTGCGGATCAAGCGGCGACCGCCACCATGCGGGGG 307  
 356 ACCCGCGCGCGCGCGCGGACCTGTGTGAGCAAGCGCGCTTACCGTTCGAGCGCTTGC 415  
 308 AAGTGGCGGACCGGATGAATCTCTCGGCGAGCAACATTATATCCACCGCTGG 367  
 416 CCGACTACCTGAGCGGGTTCGGGCGCCAGCGCGCGCGCGCTGATGATGCTGG 475  
 368 AGGCTTATGCGCCATGCGGCTGGAAGCGCGCGGACGTCGTGAATGTGCGACGCTTATCG 427  
 476 GCCATTCAACGCTGCGCGCGCGGCTGATGCGGAGTTCGACGCGCGCGCGACGAGG 535  
 428 GTGACACCGCGGCGGTATTAACCATATGACATCTGTTGCTCGGCGCAAGAGCG 487  
 536 AATCGCGGCGCATGCGGACCTGCGGAGCAAGCAATGCGGCGCGCGCATGCGCAT 595  
 488 AATTCGCGGTATGCGCGCTGACGCTACGCGATGCGTGGCGCGGAGGAGCTTAGGCTCA 547  
 596 CGACGCGCGCTTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 655  
 548 GTACCGCGGCTGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 607  
 656 GCGCGCGCTGAGCGCGCATGCGCGCATGCGCGCATGCGCGCATGCGCGCATGCGCG 715  
 608 CGGAGAGCTGCGCGCGCGCGGAGGGGCTATACACACCATTTACTTACAGATTGAGC 667  
 716 ACATGCTGCGCGCGCGCGCGGAGAACTTCCGATGCGCGCGCGCGCGCGCGCGCG 775  
 668 CATTTCTGAGCGCGCGCGCGGAGAGCTCCGATTTGCGGTACGCGCAAGTGGCGGTG 727  
 776 TGATCTCGACACCAAGGTATGAGCGCAACCAATTTGCGCGCGCGCGCGCGCGCG 835  
 728 TGGTTTGCACACAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787  
 836 CGCTATGAGGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895  
 788 CTTTTTTGATGATGAGTGGCGCGCAGCAGAGATATGCGCTGACATGTTATCTTCG 847  
 896 CGGCTTCACCATGCTCAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 955  
 848 CAGGTATGACAGCGGTGATATGAACAGGTACCGAGAGATTATATACGATGACCT 907  
 956 GGTGCAAGCCTTCCCGCAACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015

119 GGTCCGAGCG 967  
 1016 GCAATCCAGTATGACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075  
 968 AGGTAGCGCTGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1027  
 1076 TGAAGCAACCGAGTGCAGAGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCG 1135  
 1028 TGAATGACAGAGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087  
 1136 GCTTGGCG 1195  
 1088 GGTGCGCGAGATCCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1147  
 1196 GGCATATGCG 1255  
 1148 GCCACTATGCG 1207  
 1256 GCTGACCG 1315  
 1208 GGTGTGCG 1267  
 1316 CCGACTGCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1375  
 1268 CCGACTGCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327  
 1376 CCGAGCG 1426  
 1328 AACGACCG 1378

# RESULT 3 US-10-156-761-3269

Sequence 3269, Application us/10156761  
 GENERAL INFORMATION:  
 APPLICANT: OMEGA, SATOSHI  
 APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIMAWA, JUN  
 APPLICANT: HORIKAWA, HIROSHI  
 APPLICANT: SHIBA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHIYUKI  
 APPLICANT: HATTORI, MASAHIRA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-262  
 CURRENT APPLICATION NUMBER: US/10/156/761  
 PRIOR FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 3269  
 LENGTH: 1590  
 TYPE: DNA  
 ORGANISM: Streptomyces avermitilis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)..(1590)  
 US-10-156-761-3269

Query Match 9.2% Score 161.2; DB 8; Length 1590;  
 Best Local Similarity 50.3%; Pred. No. 2e-16;  
 Matches 586; Conservative 0; Mismatches 538; Indels 42; Gaps 6;

59 TCGACTGCTGCTGCGGGGCGACCTCATGACGCGGCAACACCCGGGGGGCG 118  
 2 TGGATCTGTATCTGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 61  
 119 CCGACTGGGGGGGGGGGGGGGCGCATGCGCCGCTGCGGATCTGTGAGCGCGCG 175  
 62 CCGAGCTGCTATCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 121

```

OY 176 CCGCGACACCCGGGTCAGCTGTGGGCTGTGGTGGCGCCCGGCTTCATCGACTCC 235
DB 122 GCGGACGGCGGGTCTCTGACACCGCGGCTTGGCCCTTGGCCCGCTTCAGCATGC 181
OY 236 ACACCCAGACGACAACTCTGCTCAGCGCTGCGACATGACGCCCAAGATCTCGAGG 295
DB 182 AGCGCCACAGGACCTCCCTGCTGAGGAGACCGGACACAGTGGCCAAAGCGCGCCAGG 241
OY 296 GCGTCACACGAGTGTCTACGCGGCAATTGGGCTATGAGCTTGGCGCGCTGCGACGCCA 355
DB 242 GGGTACACCTCTGAACTGTGCGGCGAGAGCGGGCTCTCTGACGACCGCTGACGACCGTA 301
OY 356 ACCGCGCGG-----CCCGCTGAGACCTGCTGAGAGAGCGGCTT 397
DB 302 CGCTCGCGGAGTCCCGAGCGGCGCATACACCGCTGGAACGGCCAGCGAGCATCTGACT 361
OY 398 ACCGTTTCAGACGCTTGGCGGCTGCTGAGCGGCTTGGCGGCGCGCGCGCTGCA 457
DB 362 TCACCTGCGGCTGCGTGGCGGCTACCTGGAC---CGGCTGGACCAAGGCAATCGCGGTCA 418
OY 458 ACGCCGCTGTATGTGGGCGCATCAACGCTGGCGCGCGCGGCTGATCCGGAATTGAGC 517
DB 419 ACGGGCGTATCTATCTATCCCGAGGCGACCGCTGCTATGCTCGCGCTGGGAGGAGC 478
OY 518 GCGCGCGACCGAGAGGAAATCGCGGCGCATGCGGAGCTGGCGGAGGAGGAGGAGGAGG 577
DB 479 GCGCGGCGAGCGCGCGGAGCTGAGACCGGATCGCGGCGCTGTGGCGGAGGAGGAGG 538
OY 578 GCGCGCGCATGCGGATTCAGACGCGGCGCTTACCGCGCGCGCGCGCGCGCGCGCGC 637
DB 539 AGGCGCGGCTGCGGCTGCTCTCGGCGCTTACCTACACCGCGCGAGTACGCGGAGGAGC 598
OY 638 AAGAGATCATGAGTGTGCGGCGCGGCTGAGCGCGCATGCGGCGCATCTAGCGCACCA 657
DB 599 CCGAGCTACACGAACTGTGCGGCTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 658
OY 698 TGGCGGACGAGGAGGAGACATGCTGCGCGCGCTGAGGAGGAGGAGGAGGAGGAGG 757
DB 659 ACCGCTGTGAGGCGGCGCGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 718
OY 758 AGCTGAGCTGCGGCTGTGATCTGCGACACAGGCTGATGCGGCGCGCGCGCGCGCGC 817
DB 719 AGCGCGGCTGCGGCTGCTCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 778
OY 818 GCTGCGGAGAGAGCGTGCCTGATGAGGCGCGCATGGC---GCGCGAGGAGGAGTGGC 874
DB 779 GGGCACCGGAGCTGTGGCGCTGCTGAGGCGCGCTGCGCGAGGAGGCGCGCATACCC 838
OY 875 TGGAGCGGATCCCTGAGTGGCGGCTGCGCATGCTCAAGGAGGAGGAGGAGGAGGAGG 934
DB 839 TCGACACCTATCTGACACCGCGGCTGCGACACGCTGTG-----GCGATGCTGC 889
OY 935 CCGGAGGACCATCATCACTGTGAGAGCGCTTCCCGCAATGAGCGGCGCGCGCGC 994
DB 890 CGAGCTGGGCGGCGGAGGCGCGCGCGGAGCGGCTGCTGCTGCGGAGGAGGAGGAGC 948
OY 995 AFGAAGTGGCGGCGGAGCGCGCAATCCAAATGAGTACGAGTGTGCGCGGAGTGCAGC 1054
DB 949 -----GTGCGGAGGAGGAGTCCGCGAGCGCATGAGTGTGCGCGGAGGAGTGCAGC 1003
OY 1055 CCGGCGCATCTACTTCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1114
DB 1004 GCGTCCCATGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1063
OY 1115 CGACCATGATGCGGCTGCGGAGGCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1174
DB 1064 CCTAGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1123
OY 1175 GAGCTTCCCGCGGCTGTGGGAGC 1200
DB 1124 CCGCGCGCGGCTGCTCTGAGCAGC 1149

```

```

RESULT 4
US-10-156-761-3783
; Sequence 3783, Application US/10156761
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3783
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1728)
US-10-156-761-3783

Query Match
Best Local Similarity 55.6%; Score 118.4; DB 8; Length 1728;
Matches 345; Conservative 0; Mismatches 221; Indels 54; Gaps 4;

OY 59 TCGACCTGTGCTCGCGGCGGCGGACCGCTATGAGCGGAGCAACCCCGGCGCGCG 118
DB 5 TCGACCACTCATAGAGGCGGCGGACCGCTGTGAGCGGAGCGGCGCGCGGTACACCG 64
OY 119 CCGACCTGGCGGTGCGGCGGCGGACCGCATGCGCGCATGCTGCGAGCGCGCG 178
DB 65 CCGAGTGGGATGATGAGGAGGCGGCGGCTGATGCGCGCGGTGCGGACCGGAGCG 124
OY 179 GCGACACCGGCGTCAAGCTGTGCGGCGGCTGTGCGCGCGCGGCTTCATGCTGAGCA 238
DB 125 GG---GCGGCGAGGAGCGGCGGCGGCTGCTGCGCGCGCGCGCGCGCGCGCGC 181
OY 239 CCGACGAGACAACTACTGTCTCAGCGCGTGGGAGATGAGCGGCGGCGGCGGCGG 298
DB 182 CCGACTTACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241
OY 299 TCACACGCTGTGTCAGGCGGCAATTGCGCATGAGCTGCGCGCGCGCGCGCGGCG 358
DB 242 TGACGACGCTGCGGCGGCGGCAACTGCGGCTTACCGCTGCGCGCGGCGGCGGCG 301
OY 359 GCGCGCGC-----CCCTGAGC 376
DB 302 CCGAGAGCGGAGTACACCGCGCGGAGATGCTCCAAAGTGGAGGAGATGCTCCGTG 361
OY 377 TGTGAGAGAGGCGGCTTACCGCTTTCAGAGCGGCTTGGCGGAGTACGAGCGGCT 436
DB 362 CCGTGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 421
OY 437 GGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 496
DB 422 AGGCG---CGAGTGGCGGTCACGCGGCTTTCATGAGTGGGCGGCGGCGGCGG 478
OY 497 CGGTACCGGCGGATTCAGAGCGGCG-----CCACGAGAGGAAATGCGGCGCAT 550
DB 479 ACCTTATGGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 538
OY 551 GGAACCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 610
DB 539 TGGGCTCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 598

```



RESULT 6  
 PCT-US02-37547-56/C  
 ; Sequence 56, Application PC/TUS0237547  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Floss, Heinz  
 ; APPLICANT: Yu, Tin-Wei  
 ; APPLICANT: Leister, Eckard  
 ; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent  
 ; FILE REFERENCE: DWASH-06712  
 ; CURRENT APPLICATION NUMBER: PCT/US02/37547  
 ; NUMBER OF SEQ ID NOS: 57  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 56  
 ; LENGTH: 82746  
 ; TYPE: DNA  
 ; ORGANISM: Actinosynnema pretiosum  
 PCT-US02-37547-56

Query Match Best Local Similarity 44.28; Score 116; DB 1; Length 82746;  
 Matches 758; Conservative 0; Mismatches 935; Indels 21; Gaps 6;

QY 56 CCTTCGACCTGCTGCTCGCGCGGACACCTCATCGACGGCAGCAACCCCGGGGCGC 115  
 DB 73869 CCGACACACCTGTCACCGCGCGGGCTGCCCGCCCTGAGCGTCTCTGTCG 73810  
 QY 116 GCGCCGACCTGGGCGCTGCGCGGACCGCATCGCCCGCATCGCGATCTGTGAGCGCG 175  
 DB 73809 CCGCGCGCGACACCGCGCTGCGCGGACCGCATCGCCCGCTGCGACGGCTGCGCG 73750  
 QY 176 CCGCGACACCGCGGCTGACGCTGCGGGCTGCTGCGCGCGCGCGCGCTTCATGATCGC 235  
 DB 73749 CCGCGACCGCGCGGAGAGACGGGACCAACCGCGCGCTGCCCGCGCGACGTCCCG 73690  
 QY 236 ACACCCACGACGACAACTATGCTGCTGACGCGTGCACATGAGCGCCCAAGATCTCGAG 295  
 DB 73689 ACCCGCGCTGCG 73630  
 QY 296 GCGTCAACAGGCTGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 355  
 DB 73629 GCGCGACGCTGCG 73576  
 QY 356 ACCCG 415  
 DB 73575 TCG 73516  
 QY 416 CCGACCTGCTGACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 475  
 DB 73515 CCGACGCGCTGCG 73456  
 QY 476 GCGATTCAACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 535  
 DB 73455 GCGACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 73396  
 QY 536 AAATCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 595  
 DB 73395 CCGCGGAGACG 73336  
 QY 596 CGACCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 655  
 DB 73335 CCGTACGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 73276  
 QY 656 GCGCGCGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 715  
 DB 73275 CCGCGACCTGCG 73216  
 QY 716 ACATGTCG 772  
 DB 73215 TCGCGGAGCG 73156  
 QY 773 TGGTATCTCGACACCAAGGTCATGGCGCGCGCGCGCGCGCGCGCGCGCGCG 832

DB 73155 CCGAGCG 72096  
 QY 833 TCGCGCTGATGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 892  
 DB 73095 AACCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 73036  
 QY 893 TGGCGCGCGCTGCG 952  
 DB 73035 GCG 72976  
 QY 953 CCGTGTGCAAGCG 1009  
 DB 72975 CCGTGTGCG 72916  
 QY 1010 AGCGCGCGCAATCGAAGTACGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCG 1069  
 DB 72915 GCGACG 72856  
 QY 1070 TCATGATGACGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1129  
 DB 72855 TCGCGCTGCG 72796  
 QY 1130 CCGACGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1183  
 DB 72795 GCGACGCTACCG 72736  
 QY 1184 CCGCGGCTGCG 1243  
 DB 72735 TCGACG 72676  
 QY 1244 GGAAGATGACCG 1303  
 DB 72675 TCGTGTGCG 72616  
 QY 1304 CCGGCTGCTGCG 1363  
 DB 72615 GCG 72556  
 QY 1364 TCGAACACCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1423  
 DB 72555 TGGTGTTC---GACCAAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 72499  
 QY 1424 TCGCGCAAGAGCG 1483  
 DB 72498 CCGTGTGCG 72439  
 QY 1484 CCGTGTGCG 1543  
 DB 72438 CCGCGGAGCG 72379  
 QY 1544 CAACCGTGTGACGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1603  
 DB 72378 TCG 72319  
 QY 1604 ATATCTGTGCG 1663  
 DB 72318 CCGTGTGCG 72259  
 QY 1664 TGACCGTGTGACG 1723  
 DB 72258 CCG 72199  
 QY 1724 AGCTGTGCG 1757  
 DB 72198 CCGTGTGCG 72165

RESULT 7  
 PCT-US02-37547-4  
 ; Sequence 4, Application PC/TUS0237547  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Floss, Heinz

APPLICANT: Yu, Tin-Wei  
 APPLICANT: Leister, Eckard  
 TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent  
 FILE REFERENCE: UMASH-06712  
 CURRENT APPLICATION NUMBER: PCT/US02/37547  
 CURRENT FILING DATE: 2002-11-21  
 NUMBER OF SEQ ID NOS: 57  
 SOFTWARE: Patent version 3.0  
 SEQ ID NO: 4  
 LENGTH: 9975  
 TYPE: DNA  
 ORGANISM: Actinosynema pretiosum  
 PCT-US02-37547-4

Query Match 5.48; Score 95.4; DB 1; Length 9975;  
 Best Local Similarity 45.48; Pred. No. 3.4e-06;  
 Matches 474; Conservative 0; Mismatches 556; Indels 15; Gaps 3;

QY 414 GCGGCACTACCTGACCGCTGTCGCGGCGACCGCGCGCTCAACGCGCGCTGATGT 473  
 DB 3717 GCGGAGGCGCGCGGACCGCGCGCGCGCGCGCGCGCTGACGCGCTGCGCTC 3776  
 QY 474 GCGCATTCACGCTGCGCGCGCGCGCGCGCTGACGCGCGCGCGCGCGCGCG 533  
 DB 3777 GCGCGAGCG 3836  
 QY 534 GGAATCG 593  
 DB 3837 GCACTGATACCG 3896  
 QY 594 TTGACCG 3897  
 DB 3897 GGTACCG 3956  
 QY 654 GTGCG 713  
 DB 3957 CGAGGCG 4016  
 QY 714 GCACATCG 773  
 DB 4017 CTTGTCACCG 4076  
 QY 774 GGTATCTGCAACACAGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 833  
 DB 4077 GCG 4136  
 QY 834 GCG 893  
 DB 4137 GCG 4196  
 QY 894 GCG 953  
 DB 4197 GCAACCG 4253  
 QY 954 CTGATGCAAGCGCTTCCCGCACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1013  
 DB 4254 GCG 4312  
 QY 1014 GCGCAATCAAGTACGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1073  
 DB 4313 -----ACCTGACGAGTACGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4364  
 QY 1074 GATGAGCAACCGGACG 1133  
 DB 4365 CTCG 4424  
 QY 1134 GCG 1193  
 DB 4425 CGGCG 4484  
 QY 1194 GGGGCACTATGCG 1253

DB 4485 GTGGCGGAGGCGACCGCGCGCTACCGCGCTCCCTGTCGCGGAGCGGAGCGCATGCG 4544  
 QY 1254 GCGCTGACCGCGCGCGCGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1313  
 DB 4545 GCGCTGCG 4604  
 QY 1314 GCG 1373  
 DB 4605 GCG 4661  
 QY 1374 TACGAGCG 1433  
 DB 4662 GCG 1496  
 QY 1434 GCG 1458  
 DB 4722 GCG 4746

# RESULT 8 US-10-156-761-6189

Sequence 6189, Application US/10156761

## GENERAL INFORMATION:

APPLICANT: OMDRA, SATOSHI  
 APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIMURA, JUN  
 APPLICANT: HORIKAWA, HIROSHI  
 APPLICANT: SHIBA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHIYUKI  
 APPLICANT: HATTORI, MASAHIRA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-262  
 CURRENT APPLICATION NUMBER: US/10/156,761  
 PRIOR FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 6189  
 LENGTH: 2538  
 TYPE: DNA  
 ORGANISM: Streptomyces avermitilis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)..(2538)  
 US-10-156-761-6189

Query Match 5.28; Score 91; DB 8; Length 2538;

Best Local Similarity 43.58; Pred. No. 1.6e-05;

Matches 568; Conservative 0; Mismatches 730; Indels 9; Gaps 3;

QY 31 TCATGTCCTCAATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCC 90  
 DB 1108 TCACACCAACCG 1167  
 QY 91 GAGCGAGCAACCG 150  
 DB 1168 GAGCGATCTCTACGCGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1227  
 QY 151 GCATGCGGATCTGTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 210  
 DB 1228 GAAAGAGCTCTCTGCG 1287  
 QY 211 GTGCG 270  
 DB 1288 ACCCTGCGGACTTATGACATCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1347  
 QY 271 GACATGACCGCAAGATCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 330  
 DB 1348 GCGGACTCTGCG 1407  
 QY 331 AGCTGCG 387



Db 1408 CCCGAGGACCTGCGCCCTGACCTCGAGCGGAGAGCCGCCGCCACCGCATCACC 1467  
 QY 388 GCGGCGCTTACCTGTTAGAGCGCTTCCCGACTCTGAGAGCGCTTCCGGGACGCG 447  
 Db 1468 GTGGGCTCTAGACCGGAGACTTCAACAGAGGCGCCCTCACCTCGCGCGCCCTC 1527  
 QY 448 GCGGCGCTTACAGCGCGCTTATAGTGGGCCATTCAAGCGCGCGCGCGCTATGCGG 507  
 Db 1528 GAATCGACATCCCGACAGACCCCGCGGACCAAGCCCTCGGCAAGCGCGCGCCCTG 1587  
 QY 508 GACTTGCAGCGCGCGCGGACAGAGAGAAATCGGCGCAATGCGGAGACTGCGCGAGGA 567  
 Db 1588 CTCTCTCTCAAGAGAGCGGAGACTTCACTACGCGCAAGTCCGCTTCAACCCCGATCTTC 1647  
 QY 568 GCGATGCGCAAGCGCGCGCTATGCGGATTTGAGACGCGCGCTTCTACCGCGCGCGCG 626  
 Db 1648 ACCACGCTCGCGCGCGCGCGCTGTCGCGCGCTTCCGACCGCGCGCGCGCTGTCG 1707  
 QY 627 -GCGGACCAAGAGATCATGAGTGTGCGCGCGCTGCGGAGCGGAGTGGCGGCTATC 684  
 Db 1708 AAGCGCTTGGCAAGCGCGCGCTGACGAGCGCATCTCCCGCGACCGCGCTTACCTGCA 1767  
 QY 685 TACGCAACCAATGCGCGAGAGAGAGGAGAGAGCATGCTGCGCGCGCTGAGAGAACTTC 744  
 Db 1768 GCG 1827  
 QY 745 CGCATGCG 804  
 Db 1828 GCGACACCAAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1887  
 QY 805 CCCAATTTGCG 1887  
 Db 1888 ACCCTACCGAGACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 864  
 QY 865 GACGCTGCTGAGAGCGCTATCTCTACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1947  
 Db 1948 CTGGCGCTGACCG 924  
 QY 925 GTGCTGCTGCG 2007  
 Db 2008 AGCTGCTCTTCCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 984  
 QY 985 GCGGAGCTGATGATGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2067  
 Db 2068 GCGATCTGCG 2127  
 QY 1045 CTGAGCG 2127  
 Db 2128 CTGCTCCAGAGCG 2187  
 QY 1105 GCGTTCG 2187  
 Db 2188 CCGGACCGCGAGTCCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2247  
 QY 1162 CCGCGCTTGGGCG 2247  
 Db 2248 AACTACTCTTACCG 1221  
 QY 1222 TTCCGCGTGGAGAGCGCGTATGAGATGAGCGCGCTGAGCGCGCGCGCGCGCGCG 2307  
 Db 2308 CCGCGCTTACGTCAGCGCGTACTGAGCGAGCGCGTCCGCGCGCGCGCGCGCGCG 1281  
 QY 1282 GCGGCG 2367  
 Db 2368 GCGATCG 2414

RESULT 9  
 PCT-US02-37547-3  
 ; sequence 3, Application PC/TUS0237547  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Floss, Heinz

; APPLICANT: Yu, Jin-Wein  
 ; APPLICANT: Leister, Eckard  
 ; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor  
 ; FILE REFERENCE: UMASH-06712  
 ; CURRENT FILING DATE: 2002-11-21  
 ; NUMBER OF SEQ ID NOS: 57  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO: 3  
 ; LENGTH: 4770  
 ; TYPE: DNA  
 ; ORGANISM: Actinosynnema pretiosum  
 ; PCT-US02-37547-3

Query Match 5.1%; Score 90.2; DB 1; Length 4770;  
 Best Local Similarity 42.9%; Pred. No. 2.2e-05;  
 Matches 562; Conservative 0; Mismatches 743; Indels 6; Gaps 2;

QY 111 GCGGCG 170  
 Db 3204 GCGGCG 3263  
 QY 171 GCGGCG 230  
 Db 3264 CCG 3323  
 QY 231 CTGCGCACCGAG 290  
 Db 3324 GAG 3383  
 QY 291 GAG 350  
 Db 3384 GC---GCGGCG 3440  
 QY 351 GCGGCG 410  
 Db 3441 CACCG 3500  
 QY 411 CTTCGCGCGAGTACTGAGAGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 470  
 Db 3501 CGAGCGAGTCTGCTCTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3560  
 QY 471 GGTGGCGCATTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 530  
 Db 3561 GCGGCG 3620  
 QY 531 CGAGGAAATCG 587  
 Db 3621 GCG 3680  
 QY 588 GCGGATTCGAGCG 647  
 Db 3681 GGTGACG 3740  
 QY 648 CGAGGTGCG 707  
 Db 3741 CCGCGCGGTTCAGCG 3800  
 QY 708 AGGAG 767  
 Db 3801 GCGGCG 3860  
 QY 768 GCGGAGTGTATCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 827  
 Db 3861 CAGGTGCG 3920  
 QY 828 GAGCGTGCCTGATGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 887  
 Db 3921 GAGAGCG 3980  
 QY 888 CTACGTGCG 947

Db 3981 GCGACCCCGGACATCGCCCGCCAGCTCGCCAGGCTGCGCTGCGCTGATGCCCCGGA 4040  
 QY 948 CATCACTGTGTGCAAGCCCTCCCGGAACTGAGCGGGCCGCACTGATGATGCGCGC 1007  
 Db 4041 CCGTGGCGGTGCTCGGCGTCTGAGCGCCGCTCGCCGCAAGAGACACCCAGCTGTGAC 4100  
 QY 1008 CGAGCGCGGCAATCCAAATGACAGTGTGCGCGAGCTGACCGCGCGCGGCGCATCTA 1067  
 Db 4101 CGACGTGCGTGGAGGGGTTCGCGCCCGGTTCACCCGCTCGCCCGGAGCCCTGCT 4160  
 QY 1068 CTTCATGATGAGCAACCCGAGAGTGCAGCGCATCTGCGCTGCGCGCCGAGCATATGCG 1127  
 Db 4161 GCGGACCTGCGAGGCGCGCGCCCGCAACCGGTGCGCCGCAACCGCGAGCCGAGCCG 4220  
 QY 1128 CTCCGACGCGCTGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1187  
 Db 4221 GCTGGCGGCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4280  
 QY 1188 GGTGCTGGGCGCATATGAGCGCGCGAGCTGCGCTGCTTCCGCTGAGAGAGCGGTATGGA 1247  
 Db 4281 GGTGCG 4340  
 QY 1248 GATGACGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1307  
 Db 4341 GGTGCGCGCTGCGCGAGCTGCGCGGTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCG 4400  
 QY 1308 GTACTTGGCG 1367  
 Db 4401 CTGGCG 4460  
 QY 1368 ACACCCCTACGAGCG 1418  
 Db 4461 GCGCGTCCGCGAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4511

RESULT 10  
 PCT-US02-37547-1  
 ; Sequence 1, Application PC/TUS0237547  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Floss, Heinz  
 ; APPLICANT: Yu, Tin-Wei  
 ; APPLICANT: Leisner, Eckard  
 ; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent  
 ; FILE REFERENCE: UMASH-06712  
 ; CURRENT APPLICATION NUMBER: PCT/US02/37547  
 ; NUMBER OF SEQ ID NOS: 57  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 14055  
 ; TYPE: DNA  
 ; ORGANISM: Actinosynema pretiosum  
 PCT-US02-37547-1

Query Match 5.0%; Score 88.4; DB 1; Length 14055;  
 Best Local Similarity 44.4%; Pred. No. 4.1e-05;  
 Matches 460; Conservative 0; Mismatches 561; Indels 15; Gaps 2;

QY 334 CGGATCAAGCTCG 383  
 Db 10878 CGACACCGGCTACG 10937  
 QY 384 CGAAGCGGCTTACGCTTTCGAGCGCGCTGCGCGACTACGAGAGCGGTGCGCGCGCG 443  
 Db 10938 CGACCGGCTGCG 10997  
 QY 444 GCG 503  
 Db 10998 GCG 11097  
 QY 504 GCGCGACTTGCAGCG 563

Db 11058 GCGCCCTATGCGCGAGCTGCCCCCGCGGTATGATGTCGGGTCCGCGCGCGCGCGAGCA 11117  
 QY 564 GGAAGCATGCGCGAGCG 623  
 Db 11118 GCG 11177  
 QY 624 CCG 683  
 Db 11178 GGTGCTACCGCGCGA-----CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11228  
 QY 684 CTAGCG 743  
 Db 11229 GCG 11288  
 QY 744 CCGCATGCG 803  
 Db 11289 GCG 11348  
 QY 804 GCGCAATTTGCG 857  
 Db 11349 CACACCG 11408  
 QY 858 GCGCGAGAGCTGCG 917  
 Db 11409 GGTGCG 11468  
 QY 918 GAGCG 977  
 Db 11469 GGTGCG 11528  
 QY 978 GAGCG 1037  
 Db 11529 GCTGCG 11588  
 QY 1038 GCGCGAGCTGCG 1097  
 Db 11589 GCG 11648  
 QY 1098 CATCTGCG 1157  
 Db 11649 CCG 11708  
 QY 1158 GATCG 1217  
 Db 11709 GCG 11768  
 QY 1218 CCGTTCG 1277  
 Db 11769 GATCG 11828  
 QY 1278 CCGTGGCG 1337  
 Db 11829 GCTGCG 11888  
 QY 1338 GCG 1353  
 Db 11889 GGTGCG 11904

RESULT 11  
 PCT-US02-37547-2  
 ; Sequence 2, Application PC/TUS0237547  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Floss, Heinz  
 ; APPLICANT: Yu, Tin-Wei  
 ; APPLICANT: Leisner, Eckard  
 ; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent  
 ; FILE REFERENCE: UMASH-06712  
 ; CURRENT APPLICATION NUMBER: PCT/US02/37547  
 ; NUMBER OF SEQ ID NOS: 57  
 ; SOFTWARE: PatentIn version 3.0



OY 295 GGGTCACACAGGAGGTCACAGGCAATTGGGCATCATAGCCCTGGCGCCGCTGCGCCAGCC 354  
 DB 1414 GAGAGAGGAGGTCGAGAGGCGCTGTGCGACACCTTCGCCCGGAGGCGGCGCCAGCCCGAG 1473  
 OY 355 AACCCCGCCCGCCCGCTGAGACCTGCTGAGAGGAGGCGCTTACCTTTCAGAGCGCTC 414  
 DB 1474 GCGCCGCGCTGCTGAGGCGCTGCGGCTGCGCTTACCTTTCAGAGGCGCGCTC 1533  
 OY 415 GCC-----GACTACCTGAGAGCGCTTCCGCGCACCGCCGCGCGCTTCAAGCGCGCTGT 468  
 DB 1534 TCCCGCCGCTGCGCGCTGAGTCTCCCGGGGCGCGCGCCCGAGGTCGCCGTCGCCGTC 1593  
 OY 469 ATGCTGGGCGCTTCAAGCGCTGCGCGCGCGCTGAGCGGCTTTCAGAGCGCGCGCTC 528  
 DB 1594 TCGCTGAGACGCGACCGCGCTGCGCGCGCGCTGAGCGGCTTTCAGAGCGCGCGCTC 1653  
 OY 529 GAGGAGAAATCGCGCGCTGAGCGGCTGAGCGGCTTTCAGAGCGCGCGCTC 588  
 DB 1654 GTCCAGCTACCGCTGAGCGGCTGAGCGGCTTTCAGAGCGCGCGCTC 1713  
 OY 589 GGCATTTCGACCGCGCGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCGCGCTT 648  
 DB 1714 GTGCGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1773  
 OY 649 GAGGTGTGCGCGCGCTGAGCGGCTGAGCGGCTTTCAGCGCGCGCTTTCAGCGCGCTT 708  
 DB 1774 GCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1833  
 OY 709 GCGGAGACATGTCGCGCGCTGAGCGGCTTTCAGCGCGCGCTTTCAGCGCGCTT 768  
 DB 1834 GCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1893  
 OY 769 -----CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 819  
 DB 1894 ATCCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1953  
 OY 820 TCGCGGAGACGCTGCGCGCTGAGCGGCTTTCAGCGCGCGCTTTCAGCGCGCTT 879  
 DB 1954 TCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2013  
 OY 880 GCGTATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 939  
 DB 2014 GCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2073  
 OY 940 GCGACATCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 999  
 DB 2074 CTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2130  
 OY 1000 GTCCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1059  
 DB 2131 CCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2190  
 OY 1060 GCGATCTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119  
 DB 2191 ACCCTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2250  
 OY 1120 ATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1179  
 DB 2251 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2310  
 OY 1180 TTCCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1239  
 DB 2311 GCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2370  
 OY 1240 GTATGAGAAATGAGCGCGCTGAGCGGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCG 1299  
 DB 2371 GTGACGCTACTTGGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCG 2430  
 OY 1300 CAGCGCGGCTTCTGCGCGCTGAGCGGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCG 1359  
 DB 2431 GCGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2490  
 OY 1360 ACCTTGAAACACCTTACCGAGCGCGC 1386

DB 2491 CTGCGCGGACCGCGAGTCCCGCGCGC 2517  
 RESULT 13  
 US-10-329-079-6  
 ; Sequence 6, Application US/10329079  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FARNER, Chris  
 ; APPLICANT: ZAZOPOULOS, Emmanuel  
 ; APPLICANT: STAFPA, Alfredo  
 ; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTI  
 ; FILE REFERENCE: 3002-1105  
 ; CURRENT APPLICATION NUMBER: US/10/329, 079  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 6  
 ; LENGTH: 37360  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces fradiae  
 US-10-329-079-6

Query Match 4.7%; Score 82.6; DB 9; Length 37360;  
 Best Local Similarity 43.1%; Pred. No. 0.00033;  
 Matches 580; Conservative 0; Mismatches 743; Indels 18; Gaps 3;

OY 55 CCTTTCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114  
 DB 15390 CCGGAGGCG 15449  
 OY 115 CCGCGGACCTGCG 174  
 DB 15450 GCGCGGACCTGCG 174  
 OY 175 GCGCGGACCG 234  
 DB 15510 GCGCGGCG 15569  
 OY 235 CACACCGGACGAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 294  
 DB 15570 CTGCGAGGCTTTCG 15629  
 OY 295 GCGCTACACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 354  
 DB 15630 GACGAGGCGCTGAGGCG 15689  
 OY 355 AACCG 414  
 DB 15690 GCGCGGCG 15749  
 OY 415 GCC-----GACTACCTGAGAGCGGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCGCGCGCTT 468  
 DB 15750 TCCCGCGCTGCG 15809  
 OY 469 ATGCTGAGGCAATTCAGCGGCTGCG 528  
 DB 15810 TGCCTGAGACGCAACGCGCTGCG 15869  
 OY 529 GAGGAGAAATCGCGCGCTGAGCGGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCG 588  
 DB 15870 GTCCAGCTTACCGCTGAGCGGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCG 15929  
 OY 589 GGCATTTCGACCGCGCGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCGCGCGCTT 648  
 DB 15930 GTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15989  
 OY 649 GAGGTGTGCGCGCGCTGAGCGGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCGCGCGCTTTCAGCG 708  
 DB 15990 GCGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 16049  
 OY 709 GCGGAGACATGCTGCGCGCGCTGAGGAACTTTCGCGACATGCGCGCGCGCGCGCGCGCGCGCGCGCG 768

```

Db 16050 GGGCCGGTACCGGACCGGACCGCCGCTCTGCGGGGACGCGGCTACGTC 16109
QY 769 -----CCGGTGTATCTCGACACCAAGGTCATGGGCGACGCCAATTTCGGCCG 819
Db 16110 ATCCACACCTGGGTTTCCACCGGCGGACGCCCAAGGGGTCACGTTGACACCGGGGCTG 16169
QY 820 TCGCGCGGAGCGCTCGCGCTTATGAGCGCCGCAATGGCGCGCGACGAGCTCTGCTGAC 879
Db 16170 TCGGGCTGCTCCAGGCGCCACCGCGGTCACCTTCCTCCGCACTCCCTCCGCGAGCG 16229
QY 880 GCGATCCCTACGTGGCGGGGCTCCACCATGCTCAAGACGAGCGCGTGTGCTGCGCGGCA 939
Db 16230 GGGCCCGGCGCGCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16289
QY 940 GCGACATCATACCTGTGTGACAGCCCTTCCCGCACTGAGCGGGGCGCGACCTGATGAA 999
Db 16390 CTCGCGATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16346
QY 1000 GTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1059
Db 16347 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16406
QY 1060 GCGATCTACTTCTATGATGACGACAAACCGGACGAGCTGACGCGACTGCGCGCTTCCGCGACCG 1119
Db 16407 ACCTACTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16466
QY 1120 ATGATGGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1179
Db 16467 CTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16536
QY 1180 TTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1239
Db 16527 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16586
QY 1240 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1299
Db 16587 GTGACCGTACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16646
QY 1300 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1359
Db 16647 GCGTACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16706
QY 1360 ACCCTGGAACCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16706
Db 16707 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16733

```

## RESULT 14

US-10-156-761-6160

Sequence 6160, Application US/10156761

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADATOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

PRIOR FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 1610

LENGTH: 1371

TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1371)  
US-10-156-761-6160

Query Match

Best Local Similarity 46.4%;

Matches 309; Conservative 0; Mismatches 351; Indels 6; Gaps 1;

```

QY 66 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4.7%; Score 82.4; DB 8; Length 1371;
Db 687 GCGCATGCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46.4%; Pred. No. 0.00335;
QY 126 GGGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 0; Mismatches 351; Indels 6; Gaps 1;
Db 747 CAACCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
QY 186 CCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 0; Mismatches 351; Indels 6; Gaps 1;
Db 807 CCGGCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
QY 246 CGACACTACTGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 0; Mismatches 351; Indels 6; Gaps 1;
Db 867 CCGCGACCGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
QY 306 GGTGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 0; Mismatches 351; Indels 6; Gaps 1;
Db 925 ----GCGCACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
QY 366 CCGCGTGAAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 0; Mismatches 351; Indels 6; Gaps 1;
Db 981 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
QY 426 GAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 0; Mismatches 351; Indels 6; Gaps 1;
Db 1041 CGACACAGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
QY 486 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 0; Mismatches 351; Indels 6; Gaps 1;
Db 1101 CCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
QY 546 CATGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 0; Mismatches 351; Indels 6; Gaps 1;
Db 1161 CGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
QY 606 CTCTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 0; Mismatches 351; Indels 6; Gaps 1;
Db 1221 CATGCGCGGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
QY 666 GAGCGCGCATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 0; Mismatches 351; Indels 6; Gaps 1;
Db 1281 CAAGAACCGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
QY 726 CCGCGCT 731
Db 1341 CCACCT 1346

```

## RESULT 15

US-10-314-657-1/C

Sequence 1, Application US/10314657

GENERAL INFORMATION:

APPLICANT: SHEN, Ben

APPLICANT: CHENG, YI-Qiang

APPLICANT: TANG, Gong-Li

TITLE OF INVENTION: Discrete Apyltransferases Associated with Type I Polyketide

FILE REFERENCE: 054030-0021

CURRENT APPLICATION NUMBER: US/10/314,657

PRIOR FILING DATE: 2002-12-09

PRIOR APPLICATION NUMBER: PCT/US02/08937

PRIOR FILING DATE: 2002-03-22

PRIOR APPLICATION NUMBER: US 60/278,935

PRIOR FILING DATE: 2001-03-26

NUMBER OF SEQ ID NOS: 214





GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 01:57:27 ; Search time 3551 Seconds  
(without alignments)  
12447.337 Million cell updates/sec

Title: US-10-009-782-1  
Perfect score: 1758

Sequence: 1 gaattccactgctgcgcgga.....ccctgagctacgagaagctt 1758

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

```

Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/1/pna/US06000.COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06001.COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US06002.COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US06003.COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US06004.COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US06005.COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US06006.COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US06007.COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US06008.COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US06009.COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US06010.COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US06011.COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US06012.COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US06013.COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US06014.COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US06015.COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US06016.COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US06017.COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US06018.COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US06019.COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US06020.COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US06021.COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US06022.COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US06023.COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US06024.COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US06025.COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US06026.COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US06027.COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US06028.COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US06029.COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US06030.COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US06031.COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US06032.COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US06033.COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US06034.COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US06035.COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US06036.COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US06037.COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US06038.COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US06039.COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US06040.COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US06041.COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US06042.COMB.seq:*

```

Result No.	Score	Query Match	Length	DB ID	Description
1	1758	100.0	1758	US-10-009-782-1	Sequence 1, Appl1
2	1440.6	81.9	1473	US-09-807-788-3	Sequence 3, Appl1
3	1428.6	24.4	1473	US-09-489-0399-1847	Sequence 1847, Ap
4	133.6	7.6	453	US-60-184-777-35	Sequence 35, Appl
5	88.8	5.1	5877	US-10-152-886-54	Sequence 54, Appl
6	86	4.9	88421	US-09-976-059-1	Sequence 1, Appl
7	85.4	4.9	770	US-09-739-449-925	Sequence 925, App
8	85.4	4.9	770	US-09-803-110-925	Sequence 925, App
9	84.6	4.8	109519	US-09-758-759-1	Sequence 150, App
10	82	4.8	1839	PCT-US02-08937-1	Sequence 1, Appl1
11	82	4.7	1248	US-09-836-842-7	Sequence 7, Appl1
12	81.8	4.7	1248	US-09-860-846-7	Sequence 7, Appl1
13	81.8	4.7	1248	US-09-861-289-7	Sequence 7, Appl1
14	81.8	4.7	1248	US-09-861-289-7	Sequence 7, Appl1
15	81.8	4.7	1248	US-09-861-289-7	Sequence 7, Appl1
16	81.8	4.7	12441	US-09-988-384B-3	Sequence 3, Appl1
17	81.8	4.7	13613	US-09-836-821-3	Sequence 3, Appl1
18	81.8	4.7	13613	US-09-860-846-3	Sequence 3, Appl1
19	81.8	4.7	13613	US-09-861-289-3	Sequence 3, Appl1
20	80.2	4.6	5760	US-10-152-886-14	Sequence 14, Appl
21	80	4.6	3240	US-09-887-272A-4417	Sequence 4417, Ap

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

```

c 22 79.2 4.5 1458 16 US-09-252-991A-1112 Sequence 1112, App
23 79.2 4.5 1461 16 US-09-252-991A-972 Sequence 972, App
24 79.2 4.5 1478 29 US-09-758-759-126 Sequence 126, App
25 79.2 4.5 4341 41 US-10-166-087-45 Sequence 45, App
26 79.2 4.5 4563 16 US-09-252-991A-930 Sequence 930, App
27 79.2 4.5 32539 80 US-10-166-087-1 Sequence 1, App
28 79.2 4.5 2280 80 US-60-360-039-42316 Sequence 1, App
29 79.2 4.5 3602 41 US-10-152-886-84 Sequence 84, App
30 78.8 4.5 1674 34 US-09-902-540-7458 Sequence 7458, App
31 78.8 4.5 4387 42 US-09-902-540-721 Sequence 721, App
32 78.2 4.4 3957 42 US-10-200-562-193 Sequence 193, App
33 78.2 4.4 12349 16 US-10-227-551-193 Sequence 193, App
34 78.2 4.4 12349 16 US-09-266-965-74 Sequence 74, App
35 78.2 4.4 12249 36 US-09-953-348-74 Sequence 74, App
36 78.2 4.4 12249 36 US-09-953-348-74 Sequence 74, App
37 78.2 4.4 18331 36 US-09-266-965-96 Sequence 96, App
38 78.2 4.4 18331 36 US-09-953-348-96 Sequence 96, App
39 78.2 4.4 39949 82 US-60-381-511-1 Sequence 1, App
40 78.2 4.4 48200 82 US-60-381-511-1 Sequence 1, App
41 78.2 4.4 154746 1 PCT-US01-11372-8 Sequence 2, App
42 78.2 4.4 154746 1 PCT-US01-11372-8 Sequence 2, App
43 78.2 4.4 154746 31 PCT-US01-11372-8 Sequence 8, App
44 78.2 4.4 154746 31 PCT-US01-11372-8 Sequence 8, App
c 45 4.4 154746 31 US-09-827-688-8 Sequence 8, App

```

## ALIGNMENTS

```

RESULT 1
US-10-009-782-1
Sequence 1, Application US/10009782
GENERAL INFORMATION:
APPLICANT: TAKEUCHI, Ken-ichi
APPLICANT: KOIDE, Yoshinao
APPLICANT: HIROSE, Yoshihiko
APPLICANT: MORIGUCHI, Mitsunaki
APPLICANT: ISOBUE, Kimiyasu
TITLE OF INVENTION: TRANSFORMED MICROORGANISM AND PROCESS FOR PRODUCING D-AMINOCYCLAS
FILE REFERENCE: 217301US0PT
CURRENT APPLICATION NUMBER: US/10/009,782
CURRENT FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: PCT/JP00/03932
PRIOR FILING DATE: 2000-06-15
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1758
TYPE: DNA
ORGANISM: Alcaligenes xylosoxydans subsp. xylosoxydans
FEATURE:
NAME/KEY: CDS
LOCATION: (34)..(1485)
OTHER INFORMATION:
US-10-009-782-1

```

```

Query Match
Best Local Similarity 100.0%; Score 1758; DB 38; Length 1758;
Matches 1758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GAATCCACTGATGCGGAGAGAGATTTCATGTCCTCAATCGATTCGACGCCCTTC 60
DB 1 GAATTCCTACTGATGCGGAGAGAGATTTCATGTCCTCAATCGATTCGACGCCCTTC 60
QY 61 GACCTGCTGCTGCGGGGGGACCTCATGACGCGGAGCAACACCCGCGGGGCGCGCC 120
DB 61 GACCTGCTGCTGCGGGGGGACCTCATGACGCGGAGCAACACCCGCGGGGCGCGCC 120
QY 121 GACCTGCGGCTGCGGCGGAGACCGCATGCGCGCATGCGGCGATGCTGAGACCGCGCG 180
DB 121 GACCTGCGGCTGCGGCGGAGACCGCATGCGCGCATGCTGAGACCGCGCGCG 180
QY 181 CACACCGGGGTCGACGCTGTGGGCTGTGGTGGCGCGCCGCTTCATCGACTGCAACAC 240

```

```

DB 181 CACACCGGGGTCGACGCTGTGGGCTGTGGTGGCGCGCCGCTTCATCGACTGCAACAC 240
QY 241 CACGACGACACTACCTGCTGAGCGGCTGCGCATATGACGCCCAAGATCTCGAGGGCTC 300
DB 241 CACGACGACACTACCTGCTGAGCGGCTGCGCATATGACGCCCAAGATCTCGAGGGCTC 300
QY 301 ACCACGGTGTACGGGCAATTGCGCATAGCCCTGGCGCGCTGGCCAGCCAAACCG 360
DB 301 ACCACGGTGTACGGGCAATTGCGCATAGCCCTGGCGCGCTGGCCAGCCAAACCG 360
QY 361 CCGCGCCCTGAGACCTGCTGAGCAAGGCGGCTTACCTTTCGAGCGCTTGGCGGAC 420
DB 361 CCGCGCCCTGAGACCTGCTGAGCAAGGCGGCTTACCTTTCGAGCGCTTGGCGGAC 420
QY 421 TACCTGACGCGTGGCGGCGCCAGCGCGGCTTCAACGCGCGCTTATGTTGGCCAT 480
DB 421 TACCTGACGCGTGGCGGCGCCAGCGCGGCTTCAACGCGCGCTTATGTTGGCCAT 480
QY 481 TCAACGCTGCGCGCGGCGGCTGATGCGGACTTGTAGCGCGCGCGCCAGCAAGAAATC 540
DB 481 TCAACGCTGCGCGCGGCGGCTGATGCGGACTTGTAGCGCGCGCGCCAGCAAGAAATC 540
QY 541 GCGGCGCATGCGGCGGCTGCGGCGGAGAGAGCCATGGCGCGGCGCATGGCATTTGACC 600
DB 541 GCGGCGCATGCGGCGGCTGCGGCGGAGAGAGCCATGGCGCGGCGCATGGCATTTGACC 600
QY 601 GCGGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
DB 601 GCGGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
QY 661 CCGCTGAGCGCGCATGCGGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
DB 661 CCGCTGAGCGCGCATGCGGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 721 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
DB 721 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
QY 781 TCGCACACAAAGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
DB 781 TCGCACACAAAGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
QY 841 ATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
DB 841 ATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
QY 901 TCGACATGCTCAAGCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
DB 901 TCGACATGCTCAAGCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
QY 961 AAGCCCTTCCCGAATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
DB 961 AAGCCCTTCCCGAATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
QY 1021 TCGAAGTACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
DB 1021 TCGAAGTACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
QY 1081 GAACCCGACGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
DB 1081 GAACCCGACGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
QY 1141 CCGCAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
DB 1141 CCGCAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
QY 1201 TATGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
DB 1201 TATGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
QY 1261 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
DB 1261 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320

```

Qy	211	GGGGGGGGGGCTTAACTGCACTGCGCACACCCAGACGACAACTACCTGCTCAGGGCGTGC	270
Db	184	GTCCGGCCGGGCTTCAATCGACTCGCACACCCAGACGACAACTACCTGCTCAGGGCGGC	243
Qy	271	GACATGACGGCCAAAGATCTTCGCAAGGGCGTCACACGAGGATGACAGGGCAATTGGGGATC	330
Db	244	GACATGACGGCCAAAGATCTTCGAGGGCGTCACACAGGTGTGCAGGGCAATTGGGGATC	303
Qy	331	AGCCTGGCGCCGCTGGCGCGCACGCCAACCCGGCGCCCGCTTGAGACTGTGTGAGAGGC	390
Db	304	AGCCTGGCACCGCTGGCGCGCACCCAACCCGGCGCCCGCTTGAGACTGTGTGAGAGGC	363
Qy	391	GGCTCTTACCGTTTCAGAGCGTTCGCGCACTACGTGAGAGGGTTGGGGGCAAGCCGGG	450
Db	364	GGTTCCGATACCGGTTGAGAGGCGTTCCCGCACTACGTGAGAGGGTTGGGGGCAAGCCGGG	423
Qy	451	GGCGTCAAGCGCGCCGCTGTGTGGTGCCATTCAACGCTGCGCGCGCGGCGTATGCCGAC	510
Db	424	GGCGTCAAGCGCGCCGCTGTGTGGTGCCATTGACGCTGCGCGCGCGGCGTATGCCGAC	483
Qy	511	TTGCAGCGCGCGCCACACAGAGAAATCGGGGCAATGGGGGACMGGGGCGAGAGAGCC	570
Db	484	TTGCAGCGCGCGCCACACAGAGAAATCGGGGCAATGGGGGACMGGGGCGAGAGAGCC	543
Qy	571	ATGGCCAGCGGGGCCATGTGGCATTTTCAGCGGGCGCTTCAACCGCGCGCGCGCGCGC	630
Db	544	ATGGCCAGGGGGCCATGTGGCATTTTCAGCGGGCGCTTCAACCGCGCGCGCGCGCGC	603
Qy	631	ACCACGAGAGATCATCGAGGTGTGGCGGGCGGTGAGCGCGCATGGCGGCATCTACGCC	690
Db	604	ACCACGAGAGATCATCGAGGTGTGGCGGGCGGTGAGCGCGCATGGCGGCATCTACGCC	663
Qy	691	ACCCACATCGCGCGAGGAAGCGAGCACATCGTGGCGCGCGGTGAGAGAACTTCCGATC	750
Db	664	ACCCACATCGCGCGAGGAAGCGAGCACATCGTGGCGCGCGGTGAGAGAACTTCCGATC	723
Qy	751	GGCGCGAGACTGGAAGTACCGGTGTGTATCTCGACCAAGATTCATGGGCGAGCCAAAT	810
Db	724	GGCGCGAGACTGGAAGTACCGGTGTGTATCTCGACCAAGATTCATGGGCGAGCCAAAT	783
Qy	811	TTTCGGCGCTGCGCGGAGACGCTGCCCTGTATCGAGAGCGCCGATGGCGGAGAGAGTC	870
Db	784	TTTCGGCGCTGCGCGGAGAGCCTGCCCTGTATCGAGAGCGCCGATGGCGGAGAGAGTC	843
Qy	871	TGCGTGGAGGGGTAATCCCTACGTGGCGCGGCTCCACCATGCTCAAGCAGAGACCGGTGCG	930
Db	844	TGCGTGGAGGGGTAATCCCTACGTGGCGCGGCTCCACCATGCTCAAGCAGAGACCGGTGCG	903
Qy	931	CTGGCGGAGCAGCACCATCATCACTGCTGTGAAACCCCTCCCGCAATGAGCGGGGCGAC	990
Db	904	CTGGCGGAGCAGCACCATCATCACTGCTGTGAAATCCCAATAGAGAGTGGTGGCGAGCTGCAG	963
Qy	991	CTGGATGAAAGTCGGCGCGGAGCGCGGCGGCAATCCAACTACGACGATGGTCCGAGCTGCAG	1050
Db	964	CTGGATGAAAGTCGGCGCGGAGCGCGGCGGCAATCCCAATAGAGAGTGGTGGCGAGCTGCAG	1023
Qy	1051	CCGGCGCGCGCATCTCACTCATGATGAGGAACCGGAGGTGCACAGCATCTGTGGCGTTC	1110
Db	1024	CCGGCGCGCGCATCTCACTCATGATGAGGAACCGGAGGTGCACAGCATCTGTGGCGTTC	1083
Qy	1111	GGCCCGACATGATCGGCTCCGAGCGGCGTCCGCAAGACAGCAGCGCCGATTCGGGCGCTG	1170
Db	1084	GGCCCGACATGATCGGCTCCGAGCGGCGTCCGCAAGAGAGCGCCGATTCGGGCGCTG	1143
Qy	1171	TGGGGCAACTTCCCGCGGGGTGCTGGGGCACTATGGCGCGACCTGGGCGCTTCCCGCTG	1230
Db	1144	TGGGGCAACTTCCCGCGGGGTGCTGGGGCACTATGGCGCGACCTGGGCGCTTCCCGCTG	1203
Qy	1231	GAGACGGCGGATGAGAGATGACCGCGCTGACCGCGCGCGCGCTTGGCGTGGCGGGCGC	1290
Db	1204	GAGACGGCGGATGAGAGATGACCGCGCTGACCGCGCGCGCGCTTGGCGTGGCGGGCGC	1263



[illegible][illegible]





```

OTHER INFORMATION: ORF 29; positive strandedness
NAME/KEY: misc_feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
NAME/KEY: misc_feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
NAME/KEY: misc_feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; positive strandedness
NAME/KEY: misc_feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-976-059-1

```

```

Query Match      4.9%; Score 86; DB 36; Length 88421;
Best Local Similarity 43.1%; Pred. No. 0.00037;
Matches 605; Conservative 0; Mismatches 775; Indels 24; Gaps 3;

```

```

OY 35 TGTCCATTCGATTCGACCTTCGACCTGCTGCGGCGGCGACCTCATCGACG 94
Db 46230 TGGCCCGGACAGAGTGTGTGCGAGGTTCACGGTCCGACGCGACCGTGGCAGC 46289
OY 95 GCACGACACCCCGGCGGCGCGCGACCTGCGCGCGCGACCGCATCGCCGCA 154
Db 46290 ACATCCCTGACCCCGCGCGCGCGCGCGCTGCGCGTGTGCGAGTGGCGCGCGC 46349
OY 155 TCGCGCATGTGTGAGAGCGCGCGCGCGACACCCGGGTGACGTGTGCGGCTGTG 214
Db 46350 GGTTCGAGAGCGCGGTGTGCGAGACCGCGCGCGGTGACGCTTGCACCTGCGCGGAGATCC 46409
OY 215 CGCCCGGCTTCATGACCTGCGACACCGACGACGACACTACCTGCTGAG-----GCCTC 268
Db 46410 CGCTGGGTGCGGTGTCTGCTGCGCGCGCGCGCGACGACGACGCTGCTGTGTGCTGC 46469
OY 269 GCGACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 328
Db 46470 ACCACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 46529
OY 329 TCAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 388
Db 46530 ACGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46589
OY 389 GCGGCTCTTACCGTTTCGACGCGCTTCGCGACCTGACGCGCTTCGCGCGCGCGCGCG 448
Db 46590 ACTACGCGCTCTGCGACGCTGACCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46649
OY 449 CGGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 508
Db 46650 CCGGCGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46709
OY 509 ACTTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 568
Db 46710 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46769
OY 569 CCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 628
Db 46770 TCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46829
OY 629 CCACGACCGGAGATGATGAGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 688
Db 46830 TCATGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46889
OY 689 CCACCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 748
Db 46890 TCCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46949
OY 749 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 808
Db 46950 TCTGTGTAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47009
OY 809 ATTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 868

```

```

Db 47010 TCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47069
OY 869 TCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 928
Db 47070 AGAACCTGTGTGAGAGAGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47123
OY 929 TGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 988
Db 47124 TCCAGGTATGTGTACCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47183
OY 989 ACCGTGATGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1048
Db 47184 CGATTCGCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47243
OY 1049 AGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1108
Db 47244 CCGAGACCTTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47303
OY 1109 TCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1168
Db 47304 CCGACCTTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47363
OY 1169 TGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1228
Db 47364 GCGTCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47411
OY 1229 TGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1288
Db 47412 TGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47471
OY 1289 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1348
Db 47472 CCGGATTCACCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47531
OY 1349 CCGATTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1408
Db 47532 CGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47591
OY 1409 TCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1468
Db 47592 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47645

```

```

RESULT 7
US-09-739-449-925
; Sequence 925, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 925
; LENGTH: 770
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; NAME/KEY: unsure
; LOCATION: (1)..(770)
; OTHER INFORMATION: unsure at all n locations
US-09-739-449-925

```

```

Query Match      4.9%; Score 85.4; DB 29; Length 770;
Best Local Similarity 51.0%; Pred. No. 0.00089;
Matches 214; Conservative 0; Mismatches 205; Indels 1; Gaps 1;
OY 1078 GACGACCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137
Db 35 GANCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94

```

QY 1138 CTGCGGACGAGAGAGCCGCGCATCCGCGCTGTGGGACACCTTCCGCGGGTGGGG 1197  
 Db 95 GGGGGGGGCGCGCGCGCGCGGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154  
 QY 1198 CACTATGGCGCGGACCTGGGCTGTCCCTGTGAGAGAGCGGGATATGGAATGACCGCG 1257  
 Db 155 CCGGCG 214  
 QY 1258 CTGACCGCGCGCGCTTGGGCTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1317  
 Db 215 CCGGCG 274  
 QY 1318 GACTGTGTGTTCGACCG 1377  
 Db 275 CCG 334  
 QY 1378 GAGGCG 1437  
 Db 335 GCG 394  
 QY 1438 GCGTTACCGCGCGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1497  
 Db 395 GCG 453

# RESULT 8

US-09-803-110-925  
 ; Sequence 925, Application US/09803110  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15490)D  
 ; CURRENT APPLICATION NUMBER: US/09/803,110  
 ; PRIOR FILING DATE: 2001-03-12  
 ; PRIOR APPLICATION NUMBER: US 09/739,449  
 ; PRIOR FILING DATE: 2000-12-19  
 ; PRIOR APPLICATION NUMBER: US 09/514,000  
 ; PRIOR FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: US 60/168,139  
 ; PRIOR FILING DATE: 1999-12-01  
 ; NUMBER OF SEQ ID NOS: 13351  
 ; SEQ ID NO 925  
 ; LENGTH: 770  
 ; TYPE: DNA  
 ; ORGANISM: Agrobacterium tumefaciens  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(770)  
 ; OTHER INFORMATION: unsure at all n locations  
 US-09-803-110-925

Query Match 4.9%; Score 85.4; DB 31; Length 770;  
 Best Local Similarity 51.0%; Pred. No. 0.0089;

Matches 214; Conservative 0; Mismatches 205; Indels 1; Gaps 1;

QY 1078 GACGACCGGACGCGCATCTGCGGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137  
 Db 35 GAANAAACG 94  
 QY 1138 CTGCGGACGAGAGAGCG 1197  
 Db 95 CCGGCG 154  
 QY 1198 CACTATGGCGCGGACCTGGGCTGTCCCTGTGAGAGAGCGGGATATGGAATGACCGCG 1257  
 Db 155 CCGGCG 214  
 QY 1258 CTGACCGCGCGCGCTTGGGCTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1317  
 Db 215 CCGGCG 274

QY 1318 GACTGTGTGTTCGACCG 1377  
 Db 275 CCG 334  
 QY 1378 GAGGCG 1437  
 Db 335 GCG 394  
 QY 1438 GCGTTACCGCGCGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1497  
 Db 395 GCG 453

# RESULT 9

US-09-758-759-1  
 ; Sequence 1, Application US/0958759  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hosted, Thomas J.  
 ; APPLICANT: Harg, Tim X.  
 ; APPLICANT: Horan, Ann C.  
 ; TITLE OF INVENTION: EvernInomleIn Biosynthetic Genes  
 ; FILE REFERENCE: ID0983K US  
 ; CURRENT APPLICATION NUMBER: US/09/758,759  
 ; PRIOR FILING DATE: 2001-01-11  
 ; PRIOR APPLICATION NUMBER: US 60/175,751  
 ; PRIOR FILING DATE: 2000-01-12  
 ; NUMBER OF SEQ ID NOS: 204  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 109519  
 ; TYPE: DNA  
 ; ORGANISM: Micromonospora carbonacea  
 US-09-758-759-1

Query Match 4.8%; Score 84.6; DB 29; Length 109519;  
 Best Local Similarity 45.1%; Pred. No. 0.0088;

Matches 542; Conservative 0; Mismatches 629; Indels 30; Gaps 5;

QY 117 CCGCGACCTGGGCG 176  
 Db 94498 CCGCGACCTGGGCG 94557  
 QY 177 CCGCGACCG 236  
 Db 94558 GCG 94617  
 QY 237 CACCGACGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 296  
 Db 94618 CCG 94677  
 QY 297 CGTCACACGAGTGTTCACGCGGCAATTCGCGCATACGCTGGCGCGCGCGCGCGCGCGCG 356  
 Db 94678 GGTACACG 94737  
 QY 357 CCG 404  
 Db 94738 CCG 464  
 QY 405 CGAGCGCTTCCGCGCATCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 524  
 Db 94788 CCGGCG 584  
 QY 465 CTGATGTGTGGCCATTCACGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94917  
 Db 94858 GCGCACCG 584  
 QY 525 CACGACGAGAAATCG 94917  
 Db 94918 CTTGACGCGGTACCG 94977  
 QY 585 CATCGGATTTGACCG 644  
 Db 94978 CCGGACCTGCG 95037

[illegible]

QY 1107 GTTGGCCCGACATGATGCGCTCCGACGCGCTGCGGACGACGAGGCCCGCATCCGCGC 1166  
 Db 1689 CGACCTCGACGATGCGTGTGCTGACGCGCGCGGGTGTCTCCAGCTGTGCGCCGACAGA 1748  
 QY 1167 CCTGTGGGACCTTCCCGGGGTGTGCGGCACTATGCGCGGACCTGG 1216  
 Db 1749 CGAGTGTGCGCGCGCGGCTGTGTACCGGACGACCTGTGCTCCAGG 1798

## RESULT 11

PCT-US02-08937-1/C  
 ; Sequence 1 Application PC/TUS0208937  
 ; GENERAL INFORMATION:  
 ; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
 ; TITLE OF INVENTION: LEINAMICIN BIOSYNTHESIS GENE CLUSTER AND ITS COMPONENTS AND THEIR  
 ; FILE REFERENCE: 309T-000110PC  
 ; CURRENT APPLICATION NUMBER: PCT/US02/08937  
 ; PRIOR FILING DATE: 2001-03-22  
 ; NUMBER OF SEQ ID NOS: 222  
 ; SOFTWARE: PatentIn Version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 135638  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces atroolivaceus  
 ; PCT-US02-08937-1

Query Match 4.7% Score 82; DB 1; Length 135638;  
 Best Local Similarity 44.3%; Pred. No. 0.0014;  
 Matches 610; Conservative 0; Mismatches 750; Indels 18; Gaps 6;

QY 117 CGCCGACCTGGGCGTGCAGCGGACCGCATGCGCGCATCGCCATCTGTGGAAGCGCGC 176  
 Db 118012 CGACGAGCTCGCTCCCGACGCGGCGACCTCACCAGCGGACCGCATCTGTATGACGAG 117953  
 QY 177 CGCGACACCCGGGTGCACTGTGCGGCTGTGTGTGCTGCGCGCGCGCTTCATGACATGCA 236  
 Db 117952 CTGCGCCGACCTGCTGTACGCGGAGTGTGAGACCGCGGCTCGGAGGACATGCGCGCGG 117893  
 QY 237 CACCCAGACGAACTACTGCTGCAAGCGGCTGTGCGGACATGACGCGCAAGATCTGCGAGG 296  
 Db 117892 GCTGCGGCGCGGACAGCGGCGGACCGGATCTGCGGATCTTCTTCTGCGGATGAGCTG 117833  
 QY 297 GCTACACGAGGTGTGCAAGCGGCAATGCGGATGACGCTGCGCGCGCTGCGCGCAAGCGCA 356  
 Db 117832 GCGCGGCTACGCGGTGCTGCTACCTGCGCGGCTGCAAGCGGCGCGGCGCAAGCGCTTCA 117773  
 QY 357 CCGCGCGCGCGCGCTGACCTGCTGAGCAAGCGGCGCTTACGCTTTCGAGCGCTTCG 416  
 Db 117772 GCGCGTGGCGCTCCGCGGACGAGCTGCAAGCGGCGGCTGCAAGCGGCGGCGGCGGAGC 117716  
 QY 417 CGACTACCTGAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 476  
 Db 117715 CGGCGATCTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 117656  
 QY 477 CCATTACGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 536  
 Db 117655 GCTGCAAGGCTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 117596  
 QY 537 AATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 596  
 Db 117595 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 117536  
 QY 597 GACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 656  
 Db 117535 GCACCGAAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 117476  
 QY 657 CC---GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 713  
 Db 117475 CCGCATGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 117416

QY 714 GCACATGTGCGCGCGCTGAGAGAACCTTCCGCGATCGCGCGGCGGCGGCGGCGGCGGCGGCGG 773  
 Db 117415 GCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 117356  
 QY 774 GTGATCTCGCACACAAAGGTCAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 833  
 Db 117355 CGGCGTATGAGACACCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 117296  
 QY 834 GCGCGTATGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 893  
 Db 117295 GATGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 117239  
 QY 894 GCGCGCGTCCACCATCTGCAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 953  
 Db 117238 CACCGCGTGGCGGCTTCTTCATCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 117179  
 QY 954 CTGTGCAAGCGGCTTCCCGAAGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1013  
 Db 117178 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 117119  
 QY 1014 CGGCAATCAAGTACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1073  
 Db 117118 CTTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 117059  
 QY 1074 GATGAGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1130  
 Db 117058 CACGACGAGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 116999  
 QY 1131 CGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1187  
 Db 116998 CGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 116939  
 QY 1188 GGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1247  
 Db 116938 GGGCGGATGTGTGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 116879  
 QY 1248 GATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1307  
 Db 116878 CTTCGACCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 116822  
 QY 1308 GTACTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1367  
 Db 116821 GGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 116762  
 QY 1368 ACACCGTACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1427  
 Db 116761 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 116702  
 QY 1428 GCAGAGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1485  
 Db 116701 GCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 116644

## RESULT 12

US-09-836-821-7  
 ; Sequence 7, Application US/09836821  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and plikromycin  
 ; FILE REFERENCE: 600.438US1  
 ; CURRENT APPLICATION NUMBER: US/09/836,821  
 ; PRIOR FILING DATE: 2001-04-17  
 ; PRIOR APPLICATION NUMBER: 09/105,537  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 7  
 ; LENGTH: 1248  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae

US-09-836-821-7

Query Match 4.7%; Score 81.8; DB 32; Length 1248;  
 Best Local Similarity 44.2%; Pred. No. 0.003;  
 Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

32 CCATGTCCCAATCCGATTCGACCCCTTGACCTGTGTGCGGGGCGACCCCTCATCG 91  
 215 CCACCTGCACACCCGACGGGGCTCCAGCTCTCTCGCGACCGCGCCCTCACCGGCG 274  
 92 ACGGACGACACACCCCGGGGCGCGGCGGACCTGGGCGCGGCGGACCGGACCGCG 151  
 275 AAGTATCATGCGCTGTGATGAGCTTCCGCCGCCACCCCGACCATCTGCGCTGATGGCC 334  
 152 CCATGGGGGATCTGTGAGAGCGCGCGCGACACCCGGGTGACGTGTGCGGGCTGTGG 211  
 335 TCACCCCGGCTTTCGCGCAATTCAGACCCGACCGGCAACCTCGACCGGACAGGTGG 394  
 212 TCGCGCCCGGCTTCATCTGACATCGACACCGACGACACTACTGTCTAGGCGTCCGG 271  
 395 CCGCGCGGCTTACACCCCGGACCTTCGCGCTGTGCGCTCCTCTGCGGCGCGCCCT 454  
 272 ACATGACGCCCCAAGATCTCGACGGGCGCTACACGCTGTGTCACGGGCAATTGCGGCATCA 331  
 455 GCGCGCGCGACGAGCTGCGGAGAGTCCGCGACGACGACGGCTGCGGTGATCTTCGACG 514  
 332 GCTGCGGCGCGCTGGCGACGCAACCCCGCGCGCGCGCTGTGACCTGTGAGCAAGGCG 391  
 515 CCGCGGACGCGCTGCGCTGCGGTGCGACGGCGCGCGCGCGCGCGCGCGCGCGCG 574  
 392 GCTTTCACGTTTTCGAGCGCTTCGCGCACTTCGACGCGCTGGCGGCGCGCGCGCG 451  
 575 AGGTCTTCAAGTTTCA---GCGCACAAAGGCGCTTCACAGCTTTCGAGGCGCGCGCG 631  
 452 CCGTCAACGCGCGCTGTATGTGTGGCCATTCACGCTGCGCGCGCGCGCTTCATGCGGACT 511  
 632 TCACGAGAGAGCGGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691  
 512 TGGAGCGCGCGCGCGCGGAGGAAATCGGCGCGCATGCGGAGGAACTTCGCGGACCA 571  
 692 TCGCGGCG 751  
 572 TGGCGACG 631  
 752 TGGCGCTTACCTCTCGACGCGCTTCCGAGGTGATCGACCGGACCGGCGCGCGCG 811  
 632 CCGCGGAGAGATCATGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691  
 812 CCGCGTACCGCGGACGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 871  
 692 CCGCATGCGCGCGGACGAGCGGACGACATGCTGCGCGCGCGCGCGCGCGCGCGCG 751  
 872 ACGGCGCTCAACACACACGACGATGATGTGTGAGATCGACGAGCGGCGCGCGCG 931  
 752 GCGCGGAGCTGAGCTGCGGTGTGATCTGCGACCAAGGATCATGGGCGACGCGCAAT 811  
 932 ACCGCGACCTCTGTGATGAGGTCTGAGGCGGAAAGCGGTGACACCGCGCGCTTCT 991  
 812 TCGGCGCGCTGCGCGGAGAGGCTGCGCGCTGATGAG-----GCGCGCATGCGCGCG 862  
 992 CCGCGGCGCTGCGCGGAGCTGAGGCGGTACCGGCGGCGGACGCGCGCGCGCGCGCA 1051  
 863 AAGAGCTCTGCTGAGCGGCTATCCCTACGTGCGCGCGCTGCGCGCGCGCGCGCGCG 922  
 1052 CCGAAGCGCTGCGCGCGCGCGCGCGCTGTCTCTCTGCGCGCGCGCGCGCGCGCGCG 1111  
 933 GCGTCTGTGCGCGGAGCGACCATCATCCTGTGTGAGGCTTTCGCGGAACTGAGCG 982  
 1112 ACATCTGCGCGGCTGCGCGCGCGCGCGCTGTGCGGTCTGTGCGCGACCGCGCGCGCG 1171  
 983 GCGCGGACCTGAGTAAAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015  
 1172 CCG 1204

RESULT 13  
 US-09-860-846-7  
 ; Sequence 7, Application US/09860846

; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D. H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600,438US1  
 ; CURRENT APPLICATION NUMBER: US/09/860,846  
 ; PRIOR APPLICATION NUMBER: 2001-05-18  
 ; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 7  
 ; LENGTH: 1248

; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-860-846-7

Query Match 4.7%; Score 81.8; DB 33; Length 1248;  
 Best Local Similarity 44.2%; Pred. No. 0.003;  
 Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

32 CCATGTCCCAATCCGATTCGACCCCTTGACCTGTGTGCGGGGCGACCCCTCATCG 91  
 215 CCACCTGCACACCCGACGGGGCTCCAGCTCTCTCGCGACCGCGCCCTCACCGGCG 274  
 92 ACGGACGACACACCCCGGGGCGCGGCGGACCTGGGCGCGGCGGACCGGACCGCG 151  
 275 AAGTATCATGCGCTGTGATGAGCTTCCGCCGCCACCCCGACCATCTGCGCTGATGGCC 334  
 152 CCATGGGGGATCTGTGAGAGCGCGCGCGACACCCGGGTGACGTGTGCGGGCTGTGG 211  
 335 TCACCCCGGCTTTCGCGCAATTCAGACCCGACCGGCAACCTCGACCGGACAGGTGG 394  
 212 TCGCGCCCGGCTTCATCTGACATCGACACCGACGACGACACTACTGTCTAGGCGTCCGG 271  
 395 CCGCGCGGCTTACACCCCGGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 454  
 512 TGGAGCGCGCGCGCGCGGAGGAAATCGGCGCGCATGCGGAGGAACTTCGCGGACCA 571  
 692 TCGCGGCG 751  
 572 TGGCGACG 631  
 752 TGGCGCTTACCTCTCGACGCGCTTCCGAGGTGATCGACCGGACCGGCGCGCGCTTCT 991  
 812 TCGGCGCGCTGCGCGGAGAGGCTGCGCGCTGATGAG-----GCGCGCATGCGCGCG 862  
 992 CCGCGGCGCTGCGCGGAGCTGAGGCGGTACCGGCGGCGGACGCGCGCGCGCGCGCA 1051  
 863 AAGAGCTCTGCTGAGCGGCTATCCCTACGTGCGCGCGCTGCGCGCGCGCGCGCGCG 922  
 1052 CCGAAGCGCTGCGCGCGCGCGCGCGCTGTCTCTCTGCGCGCGCGCGCGCGCGCGCG 1111  
 933 GCGTCTGTGCGCGGAGCGACCATCATCCTGTGTGAGGCTTTCGCGGAACTGAGCG 982  
 1112 ACATCTGCGCGGCTGCGCGCGCGCGCGCTGTGCGGTCTGTGCGCGACCGCGCGCGCG 1171  
 983 GCGCGGACCTGAGTAAAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015  
 1172 CCG 1204

Query Match	4.7%	Score 81.8;	DB 37;	Length 1248;
Best Local Similarity	44.2%;	Pred. No. 0.003;		
Matches 439;	Conservative 0;	Mismatches 542;	Indels 12;	Gaps 2;



OY 32 CCATGTCCCATCCGATTCCTCCAGCCCTTGCAGCTGCTGCTGGCGGGGGGACCCCTATCG 91  
Db 215 CCACCTGCAAGCCCAAGGGCGGGGCTCCAGCTCTCGCGCAAGCGCGCTTCACCGGCG 274  
OY 92 ACGGAGCAACACCCCGGGGGCGGGCGGACCTGGGCGTGGCGGCGACCCGATCGCG 151  
Db 275 AAGTGATCATGCGCTGATGAGCTTCGGCGCCACCCCGCAGCATGCGCTGGATGGCGC 334  
OY 152 CCATGCGGATGCTGCGGAGCGGGCGGCAACCCGGGTCCAGCTGTGGGCTGGTGG 211  
Db 335 TCACCCCGGCTTCGCGCGCATGACCGGACCGGCAACCTCGACCCGGACAGAGTGG 394  
OY 212 TCGCGCCCGGCTTCATCGACTGCGACACCGAGACAACTACCTGCTCAGGCGGCGG 271  
Db 395 CCGCGCGGCTGACACCCCGGACCTGGGCGGTGGGGCTCCACTCTGGGGCGGCGCT 454  
OY 272 ACATGACGCCCAAGATCTCGCGAGGGCTACACAGGTGGTCAAGGGCAATTGGGCATCA 331  
Db 455 GCGCGCGCGACAGCTGGGAAAGGTGGCGAGACAGGCTGGGCTGTACTTTCGACG 514  
OY 332 GCTGGCGCGCGCTGGCGCGCAACCGCGCGCGCGCGCTGGACTGCTGGACGAAGCG 391  
Db 515 CCGGCGACGGCCCTGGCGCTGGCGCGCTGACGGCGCGCGCGCGCGCTGGCGAGCGCG 574  
OY 392 GCTCTTACGTTTGAAGGCTTCGCGCGACTGAGACGCTGGCGGCGCAAGCGCGCGG 451  
Db 575 AGGTCTTCAAGCTTCA--CGCACCAAGGCGGTCAAGGCTTCGAGGGCGCGCGCTCG 631  
OY 452 CCGTCAAGCGCGCGCTGATGTTGGGCGCTTCAAGCTGCGCGCGCGGTCAATGGCGACT 511  
Db 632 TCACCGAGAGCGCGACCTCGCGCGCGGATCGCGCGCTCCACAACTTCGGCTTCGAC 691  
OY 512 TGCAGCGCGCGCGCGCGCGAGGAAATCGCGCGCATGCGGAGCTGGCGCGAGAGCCA 571  
Db 692 TGGCGCGCGGAGCGCGCGCGCGGAGCAACGCGCAAGATAGCGAGGCGCGCGCGCA 751  
OY 572 TGGCGACGCGCGCGCTGCGCTTCGACCGCGCGCTTTCACCGCGCGCGCGCGCGCA 631  
Db 752 TGGGCTTCACCTCGCTGAGCGCTTCCCGAGTTCATCGACCGGACCGGCGCAACGACG 811  
OY 632 CCAACGAAAGATCATGAGGTGTGCGCGCGCTGAGCGCGCATGGCGCATTCAGCGCA 691  
Db 812 CCGCGTACCGCGAGACACTCGCGGACCTCCCGCGCGCTCTGTCGCGACACGACGCGC 871  
OY 692 CCCACATGCGCGAGAGGCGAGACATCGTGGCGCGCGCTGAGGAAACCTTCGCGACTG 751  
Db 872 ACGGCGTCAACACCAACCACTAGTGTGTCGAGATCGACGAGCGCACCAACCGCGCATCC 931  
OY 752 GCGCGAGCTGGAAGTGGCGGTGATCTCGACACCAAGTCTATGGCGCGCGCAATT 811  
Db 932 ACCGCGACCTGCTATGAGAGTCTCTGAAGGCGGAGCGTGCACACCGCGCTACTTCT 991  
OY 812 TCGGCGCGCTGCGGAGAGCGCTCGCGCTGATCGAG-----GCCGCATGGCGCGCGC 862  
Db 992 CCGCGGCGCTGCGAGAGCTGGAACCGTACCGCGCGGAGCGCACGCGCGCTCGCGACA 1051  
OY 863 AAGACGCTGCGTGAAGCGGTATCCTACGTGGCGCGCTCAACATGCTCAAGAGAGCC 922  
Db 1052 CCGAAGCGCTCGCGCGCGCGGTGCTCTCCCTGCGAGCGCGCACCGCGCATGGCGAGCAG 1111  
OY 923 GGTGTGCTGGCGCGGAGCGACCATCATCACTGTGCAAGCGCTTCGCCGAATGAGCG 982  
Db 1112 ACATTCGCGGCGGAGCGAGCTGCTGCTGCTGCGCGAGACCGCGCGGCAACTGACCG 1171  
OY 983 GCGCGACCTGATGAAGTGGCGCGGAGCGCG 1015  
Db 1172 CCGGCAACGCGGACGAGCGCGCGCGCTCG 1204

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 23:54:40 ; Search time 395 Seconds

10022.818 Million cell updates/sec

Title: US-10-009-782-1  
Perfect score: 1758

Sequence: 1 gaattccacttgatcgcgga.....ccctgaqctacgaqaagctt 1758

Scoring table: IDENTITY\_NUC

Searched: 2185239 seqs, 1125999159 residues

total number of hits satisfying chosen parameters: 4370478

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1:	/SID2/gcgdata/genseq/genseqn-emb1/NA1980.DAT.*
2:	/SID2/gcgdata/genseq/genseqn-emb1/NA1981.DAT.*
3:	/SID2/gcgdata/genseq/genseqn-emb1/NA1982.DAT.*
4:	/SID2/gcgdata/genseq/genseqn-emb1/NA1983.DAT.*
5:	/SID2/gcgdata/genseq/genseqn-emb1/NA1984.DAT.*
6:	/SID2/gcgdata/genseq/genseqn-emb1/NA1985.DAT.*
7:	/SID2/gcgdata/genseq/genseqn-emb1/NA1986.DAT.*
8:	/SID2/gcgdata/genseq/genseqn-emb1/NA1987.DAT.*
9:	/SID2/gcgdata/genseq/genseqn-emb1/NA1988.DAT.*
10:	/SID2/gcgdata/genseq/genseqn-emb1/NA1989.DAT.*
11:	/SID2/gcgdata/genseq/genseqn-emb1/NA1990.DAT.*
12:	/SID2/gcgdata/genseq/genseqn-emb1/NA1991.DAT.*
13:	/SID2/gcgdata/genseq/genseqn-emb1/NA1992.DAT.*
14:	/SID2/gcgdata/genseq/genseqn-emb1/NA1993.DAT.*
15:	/SID2/gcgdata/genseq/genseqn-emb1/NA1994.DAT.*
16:	/SID2/gcgdata/genseq/genseqn-emb1/NA1995.DAT.*
17:	/SID2/gcgdata/genseq/genseqn-emb1/NA1996.DAT.*
18:	/SID2/gcgdata/genseq/genseqn-emb1/NA1997.DAT.*
19:	/SID2/gcgdata/genseq/genseqn-emb1/NA1998.DAT.*
20:	/SID2/gcgdata/genseq/genseqn-emb1/NA1999.DAT.*
21:	/SID2/gcgdata/genseq/genseqn-emb1/NA2000.DAT.*
22:	/SID2/gcgdata/genseq/genseqn-emb1/NA2001.DAT.*
23:	/SID2/gcgdata/genseq/genseqn-emb1/NA2002.DAT.*
24:	/SID2/gcgdata/genseq/genseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1758	100.0	1758	22	AAC91797	Alcaligenes xyloso
2	1440.6	81.9	1473	21	AAH15140	DNA encoding an am
3	103	5.9	11455	20	AAK53491	Human adenosine A1
4	101.2	5.8	312	24	ABN24790	Human ORF polynuc
5	99.2	5.6	65140	22	AAD17184	Streptomyces nous
6	99.2	5.6	125401	22	AAD17186	Streptomyces nous
7	94.2	5.4	8438	15	AAQ73501	DNA encoding pseud
8	93.8	5.3	11455	20	AAK53491	Human adenosine A1
9	93.4	5.3	27541	22	AAD17185	Streptomyces nous

10	86	4.9	88421	24	AA107091
11	84.6	4.8	109519	22	AA508693
12	81.8	4.7	1248	21	AAZ87286
13	81.8	4.7	1248	24	AAAD39044
14	81.8	4.7	12441	21	AAZ87284
15	81.8	4.7	13613	21	AAZ87319
16	81.8	4.7	13613	24	AAAD39043
17	79.8	4.5	1320	13	AAQ22481
18	78.2	4.4	3957	22	AAAO9666
19	78.2	4.4	12249	21	AAAC55840
20	78.2	4.4	18331	21	AAAC55857
21	78.2	4.4	154746	24	AAAD25519
22	78.2	4.4	154746	24	AAAD25519
23	76.2	4.3	5970	21	AAAT5655
24	76.2	4.3	5970	21	AAAT56003
25	76	4.3	12001	16	AAQ676213
26	76	4.3	28598	17	AAAT06769
27	76	4.3	58857	21	AAAT58471
28	75.8	4.3	4020	18	AAAT91361
29	75.6	4.3	43280	18	AAAT80413
30	75.4	4.3	2061	10	AAAN92408
31	75	4.3	1620	13	AAQ22482
32	75	4.3	2167	13	AAQ22484
33	73.8	4.2	2668	13	AAO22485
34	73.8	4.2	15872	21	AAAT68715
35	73.8	4.2	15872	18	AAAT87283
36	73.4	4.2	4257	19	AAAT68520
37	73.4	4.2	4257	19	AAAT10362
38	73.4	4.2	5224	22	AAAC90079
39	73.2	4.2	5224	20	AAAC32022
40	72.8	4.1	11604	22	AAAD14501
41	72.8	4.1	15079	16	AAAO91580
42	72.8	4.1	15079	23	AAAD14499
43	72.4	4.1	3189	23	AAAS1474
44	72.2	4.1	13842	21	AAAT87297
45	72.2	4.1	36778	21	AAZ87318

## • ALIGNMENTS

RESULT 1	
AAC91797	
ID AAC91797	standard; DNA; 1758 BP
XX	

27-MAR-2001 (first entry)

**Alcaligenes xylosoxidans subspecies xylosoxidans D-aminoacylase DNA**

zinc-enhanced expression; D-form amino acid synthesis;

pharmaceutical manufacturing,

**Alcaligenes xylooxidans.**

WO2000078926-A1.

28-DEC-2000

15-JUN-2000; 2000WO-JP03932.

17-JUN-1999; 990P-0170555.

(ATTN: ) AWARD ENZYME INC.

YAMAGUCHI M, HIROSE I, MORIGUCHI M, ISHIBE K,

P-PSDB; AAB48975.

661 CCGCTAGCGCCGATGGCGGCATCTACGCCACCCACATGCGCGAAGCGGAGCATC 720  
|||||  
661 CCGCTAGCGCCGATGGCGGCATCTACGCCACCCACATGCGCGAAGCGGAGCATC 720

1741 CTGAGCTACGAGAACGTT 1758  
|||||  
1742 |||||||  
1743 |||||||

Db 1741 CTGAGCTACGAGAGCTT 1758

## RESULT 2

AAA15140 standard; DNA; 1473 BP.

AAA15140;

21-AUG-2000 (first entry)

DNA encoding an aminocyclase enzyme.

Aminocyclase; N-acetyl-D-tryptophan; R-N-acetyl-2-thienylalanine;  
 R-N-acetyl-4-chlorophenylalanine; D-amino acid; N-acyl amino acid;  
 pesticide; antibiotic; ss.

Alcaligenes sp.

Location/Qualifiers  
 Key 7.1461  
 FT /tag- a  
 FT /product- "aminocyclase"

MO200023598-A1.

27-APR-2000.

20-OCT-1999; 99MO-GB03458.

20-OCT-1998; 98GB-0022947.

01-APR-1999; 99GB-0007739.

(CHIR-) CHIROTECH TECHNOLOGY LTD.

Taylor StJ, Brown RC;

WPI; 2000-339697/29.

P-PSDB; AAY84943.

New isolated D-amino acylase enzyme, useful for the preparation of  
 D-amino acids for use as intermediates in the preparation of  
 pesticides, antibiotics and other pharmaceuticals

Claim 7: Page 18-20; 26pp; English.

The present sequence encodes an aminocyclase enzyme. The enzyme is  
 capable of hydrolysing N-acetyl-D-tryptophan at a substrate  
 concentration of 10 grams per litre, and exhibiting faster conversion  
 of R-N-acetyl-2-thienylalanine than of  
 R-N-acetyl-4-chlorophenylalanine. The enzyme can be used for the  
 preparation of a D-amino acid by conversion of a corresponding  
 D-N-acyl amino acid. It can also be used for resolving a racemic  
 mixture of N-acyl amino acids and deprotecting optically-enriched  
 N-acyl amino acids. The D-amino acids are useful as intermediates in  
 the production of various pesticides, antibiotics and other  
 pharmaceuticals.

Sequence 1473 BP; 227 A; 559 C; 481 G; 206 T; 0 other;

Query Match 81.9%; Score 1440.6; DB 21; Length 1473;

Best Local Similarity 99.0%; Pred. No. 2.9e-217;

Matches 1449; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 31 TCCATGTCCTCAATCCGATTCCTGACCTGCTGCTGCGGGGACCTTCATC 90  
 DB 4 TCCATGTCCTCAATCCGATTCCTGACCTGCTGCTGCGGGGACCTTCATC 63  
 QY 91 GACGCGACCAACACCCGCGGGGCGGCGGACGACCTGCGGGGACCGATCGCC 150  
 DB 64 GACGCGACCAACACCCGCGGGGCGGCGGACGACCTGCGGGGACCGATCGCC 123  
 QY 151 GCCATCGGCGATCTGCTGCGGACGCGCGGCGACACCCGCGGTCGATGTCGGGCTGGTG 210  
 DB 124 GCCATCGGCGATCTGCTGCGGACGCGCGGCGACACCCGCGGTCGATGTCGGGCTGGTG 183

QY 211 GTCCGCGCCCGGCTTCATGACTGCGACACCGACGACCACTACCTGCTCAGGCTGCG 270  
 DB 184 GTCCGCGCCCGGCTTCATGACTGCGACACCGACGACCACTACCTGCTCAGGCTGCG 243  
 QY 271 GACATGACGCCCAAGATCTGCGACGGGCTCACACGGGTGTACAGGGCAATTTGGGCTATC 330  
 DB 244 GACATGACGCCCAAGATCTGCGACGGGCTCACACGGGTGTACAGGGCAATTTGGGCTATC 303  
 QY 331 AGCGTGGCGCGGCTGCGGCGGACGCCACCCGCGCGCCCTGGAACCTGCTGGAAGAGC 390  
 DB 304 AGCGTGGCGCGGCTGCGGCGGACGCCACCCGCGCGCCCTGGAACCTGCTGGAAGAGC 363  
 QY 391 GCGCTTACGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCG 450  
 DB 364 GCGCTTACGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCGCTTTCGAGCG 423  
 QY 451 GCGCTTACGCGCGGCTGCTGATGTCGGGCTTTCGAGCGCTTTCGAGCGCGCTTTCGAGCG 510  
 DB 424 GCGCTTACGCGCGGCTGCTGATGTCGGGCTTTCGAGCGCTTTCGAGCGCGCTTTCGAGCG 483  
 QY 511 TTGCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 570  
 DB 484 TTGCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 543  
 QY 571 ATGCGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 630  
 DB 544 ATGCGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 603  
 QY 631 ACCACCGAAGATCTGCGAGGTGTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 690  
 DB 604 ACCACCGAAGATCTGCGAGGTGTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 663  
 QY 691 ACCACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 750  
 DB 664 ACCACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 723  
 QY 751 GCGCGGAGCTGAGCGTGGCGGCTGATCTGCGACCAAGGTGATGCGGCGGCGGCGGCGGCGG 810  
 DB 724 GCGCGGAGCTGAGCGTGGCGGCTGATCTGCGACCAAGGTGATGCGGCGGCGGCGGCGGCGG 783  
 QY 811 TTGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 870  
 DB 784 TTGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 843  
 QY 871 TTGCGGAGCGGCTATCCCTGACGTGGCGGCTGCGACCAAGGTGATGCGGCGGCGGCGGCGG 930  
 DB 844 TTGCGGAGCGGCTATCCCTGACGTGGCGGCTGCGACCAAGGTGATGCGGCGGCGGCGGCGG 903  
 QY 931 CTGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 990  
 DB 904 CTGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 963  
 QY 991 CTGATGAAAGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1050  
 DB 964 CTGATGAAAGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1023  
 QY 1051 CGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1110  
 DB 1024 CGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1083  
 QY 1111 GCGCGGAGCAATGATGCGGCTGCGAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1170  
 DB 1084 GCGCGGAGCAATGATGCGGCTGCGAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1143  
 QY 1171 TGCGGCGGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1230  
 DB 1144 TGCGGCGGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1203  
 QY 1231 GAGACGCGGCTATGGAAGATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1290  
 DB 1204 GAGACGCGGCTATGGAAGATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1263  
 QY 1291 GGGGAGCTGCGAGCGGCGGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1350

1007





AA017184  
 ID AA017184 standard; DNA; 65140 BP.  
 AC AA017184;  
 XX  
 DT 29-NOV-2001 (first entry)  
 XX  
 DE Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.  
 KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
 KW antifungal; antibiotic; nys1; ds.  
 OS Streptomyces noursei.  
 XX  
 FH Key location/Qualifiers  
 FT CDS complement (1..1035)  
 FT  
 FT  
 FT CDS /tag- a  
 /product- "NysD2 partial protein"  
 /note- "CDS does not include stop codon"  
 complement (1056..2576)  
 FT  
 FT /tag- b  
 /product- "NysD1 protein"  
 2806..6906  
 FT  
 FT /tag- c  
 /product- "NysA protein"  
 6952..10530  
 FT  
 FT /tag- d  
 /product- "NysB protein"  
 16550..49840  
 FT  
 FT /tag- e  
 /product- "NysC protein"  
 50260..51015  
 FT  
 FT /tag- f  
 /product- "NysE protein"  
 51405..54305  
 FT  
 FT /tag- g  
 /product- "NysR1 protein"  
 54329..57190  
 FT  
 FT /tag- h  
 /product- "NysR2 protein"  
 /note- "CDS does not include start codon"  
 57180..55963  
 FT  
 FT /tag- i  
 /product- "NysR3 protein"  
 60415..61047  
 FT  
 FT /tag- j  
 /product- "NysR4 (short) protein"  
 /note- "CDS does not include start codon"  
 61736..62497  
 FT  
 FT /tag- k  
 /product- "NysR5 protein"  
 /note- "CDS does not include start codon"  
 complement (62551..63615)  
 FT  
 FT /tag- l  
 /product- "ORF2 protein"  
 /note- "CDS does not include start codon"  
 63765..64961  
 FT  
 FT /tag- m  
 /product- "ORF1 protein"  
 CDS  
 XX  
 PN W0200159126-A2.  
 XX  
 PD 16-AUG-2001.  
 XX  
 PF 08-FEB-2001; 2001WO-GB00509.  
 XX  
 PR 08-FEB-2000; 2000GB-0002840.  
 PR 10-APR-2000; 2000GB-0008786.  
 PR 14-APR-2000; 2000GB-0009387.  
 XX  
 PA (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPSELIGE.  
 PA (SMTF) SINTEF-STIFTELSEN IND TEK FORSK.  
 PA (ALPH-) ALPHARMA AS.

PA (SIN-) SINVENT AS.  
 PA (DZIE/) DZIEGLEMSKA H.  
 PA (ZOTC/) ZOTCHEV S. B.  
 PA (SEKU/) SEKUROVA O. N.  
 PA (EJAE/) FJAEVRIK E.  
 PA (BRAU/) BRAUTASET T.  
 PA (STRO/) STROM A. R.  
 XX  
 PI Zotehev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;  
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;  
 DR MPI: 2001-557614/62.  
 DR P-PSDB: AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,  
 DR AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.  
 XX  
 PT New nystatin polyketide synthase polynucleotides and polypeptides,  
 XX useful as antibiotics and antifungals.  
 PS Claim 2; Page 116-151; 266pp; English.  
 CC  
 CC The present invention relates to the cloning and sequencing of the gene  
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme  
 CC involved in the biosynthesis of the macrolide antibiotic nystatin.  
 CC The nystatin PKS is useful as antifungal antibiotics. The present  
 CC sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.  
 XX

SO Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

Query Match 5.6%; Score 99.2; DB 22; Length 65140;  
 Best Local Similarity 43.5%; Pred. No. 1.8e-07;  
 Matches 724; Conservative 0; Mismatches 913; Indels 27; Gaps 5;

OY 62 ACCTGCTGCTGCGGGGCGACCCCTCATCGACGAGCAACACCCCGGGGCGGGCGG 121  
 DB 19698 ACAGAGGCTGCGGCGGACTGCGCTTCAGTAGCGCCCGCTTCAGAGGCTGCGGCGG 121  
 OY 122 ACCTGGGCGTGGCGGCGACCGCATGCGGCGGCGGCGGCGGCGGCGGCGG 181  
 DB 19758 CCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 181  
 OY 182 ACACCGGCGTGCAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241  
 DB 19818 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241  
 OY 242 ACAGAGCAACATCTGCTGACGCGTGGGAGATGACCGGCGGCGGCGGCGGCGG 301  
 DB 19878 ACAGAGCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 301  
 OY 302 CCACGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 361  
 DB 19938 TCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 361  
 OY 362 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 421  
 DB 19998 TCACCATGCGGCTGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 421  
 OY 422 ACCTGAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 481  
 DB 20058 CCGCGGAGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 481  
 OY 482 CAACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 541  
 DB 20118 TCCAGCTGAGTGGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 541  
 OY 542 CGGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 601  
 DB 20178 TCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 601  
 OY 602 GCGCCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20237  
 DB 20238 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 661  
 OY 662 CCGTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 721





FT /note- "derived from PRV strain Infh"  
 FT 7014..8425  
 FT misc\_feature  
 FT //tag- b  
 FT /note- "derived from PRV strain Ka"  
 FT CDS  
 FT 622..6498  
 FT //tag- c  
 FT /note- "encodes predicted amino acid sequence of ORF2"  
 FT 1..6  
 FT 7ATA\_signal  
 FT //tag- d  
 FT 34  
 FT misc\_feature  
 FT //tag- e  
 FT /note- "RNA cap site"  
 FT 8382..8387  
 FT polyA\_signal  
 FT //tag- f  
 FT 34  
 FT US5352596-A.  
 FT 04-OCT-1994.  
 FT 11-SEP-1992: 92US-0945283.  
 FT 11-SEP-1992: 92US-0945283.  
 FT 11-SEP-1992: 92US-0945283.  
 FT (USDA ) US SEC OF AGRIC.  
 FT Cheung AK, Wesley RD;  
 FT MPI: 1994-316187/39.  
 FT P-PSDB; AAR60620.  
 FT New pseudorabies virus mutants for use in vaccine - having a  
 FT deletion and/or insertion in the early protein O gene or large  
 FT latency transcript gene  
 FT  
 FT Disclosure: Column 15-30; 43bp; English.  
 FT  
 FT AA073500 shows the Pseudorabies virus (PRV) large latency transcript  
 FT (LIT). The basic sequence is derived from PRV strain Infh and PRV  
 FT strain Ka. The LIT overlaps and is transcribed in the opposite  
 FT orientation with respect to the EPO (early polypeptide 0) and the  
 FT immediately early gene (IE180). EPO is nonessential for replicatio,  
 FT LIT is the only gene expressed during PRV latency, and the IE180  
 FT gene is absolutely necessary for PRV replication. However there are  
 FT 2 copies of IE180 in the genome. It is expected that PRV lacking one  
 FT of the IE180 copies is viable. Deletions in the non-overlapping  
 FT regions of these 3 genes will generate single deletion mutants,  
 FT while deletions in overlapping regions will generate double deletion  
 FT mutants. The invention is concerned with the construction of attenuated  
 FT viruses which have a reduced ability to reactivate from latency. This  
 FT can be achieved by functionally disabling the expression of the EPO  
 FT gene, or by disrupting the synthesis of the LIT, or both. (See also  
 FT AA073501 and AAR60620-24)  
 FT  
 FT Sequence 8438 BP; 1141 A; 2916 C; 3327 G; 1054 T; 0 other;  
 FT  
 FT Query Match 5.4%; Score 94.2; DB 15; Length 8438;  
 FT Best Local Similarity 44.5%; Pred. No. 1.4e-06;  
 FT Matches 608; Conservative 0; Mismatches 738; Indels 21; Gaps 5;

Db 5427 GCCCGCCCCCGCCGAGAGAGCCGCGCTCCAGCAACAACACACCTCTGCTCATGCGCCGAC 5368  
 Oy 271 GACATGACGCCCAAGATCTCCAGGGGCTACACAGGGTGGTCAAGGGCAATGGGCATC 330  
 Db 5367 GGGCCCCCGCCGACGAGAGGGCGCGCTGCTACCCCGCTGGGGAGCCCTGGCCGCTCC 5308  
 Oy 331 AGCCTGGCGCGCTGGCGGACGCCAACCCGCCCCCTGGAGCTGCTGGAGAGAGGC 390  
 Db 5307 GACCTCGCGCGGAGAGGGCGCGCTGCTGCTGAGCGGCGCGGCGGCTCCCGGAGGGCTC 5248  
 Oy 391 GGCCTTACCTTTCGAGCGCTTCCGAGCTTACCTGAGACGGCTTGGCGGACCGCGGC 450  
 Db 5247 TGGACGAGAGAGAGCTTCCGCGAGCGCGGCGCGCTTACCGCGCGCGCGCGCGCGCTG 5188  
 Oy 451 GCGCTGCAACGCGCGCTGATGCTGCGCGCTTCAACGCTGCGCGCGCGCGCTGCGCGAC 510  
 Db 5187 CCGCTTTCATCCCGGAGATGGGGGACTCGAGAGAGGACGAGCGCGCTGCTGCGCGCTC 5128  
 Oy 511 TTGCGCGCGCGCGCCAGCGAGAGAAATCGCGCGCATGGGGGACCTGGCGGAGAGAGCC 570  
 Db 5127 ATCTACAGCGGGGCG 5068  
 Oy 571 ATGGCCAGCGCGCGCATCGCGCATTTTCAGCGCGCGCTTTCACCGCGCGCGCGCGCGCG 630  
 Db 5067 CCGGACGAGCGCTTCAACAGCTTCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5008  
 Oy 631 ACCACGAGAGATCATCATGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 650  
 Db 5007 TCTTTCATACCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4948  
 Oy 691 ACCCATGCGCGCGAGAGAGCGAGACATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750  
 Db 4947 CAGGACCGCGCTGCG 4888  
 Oy 751 G---GGCGGAGCTGAGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807  
 Db 4887 GATCGGACCGAGAGAGCTTTCATCTGCAAGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 4828  
 Oy 808 AATTTGCGCGCGCTGCGCGCGAGAGCGTGCCTGATGAGAGCGCGCGCGCGCGCGCGCGCG 867  
 Db 4827 TACCG 4768  
 Oy 868 GTTTCGCTGAGCGCGTATCTCTACG---TGCCGCGCTCAGCATGCTCAGCAGAGCCGC 924  
 Db 4767 CGGCTCGCGCGCGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4708  
 Oy 925 GTCTCTGCGCGCGAGGACGACATCATCAGTGTGCAAGCCCTTCCCGCACTGAGCGCG 984  
 Db 4707 GCTTGTCTGCTGCG 4648  
 Oy 985 CGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1044  
 Db 4647 CCG 4588  
 Oy 1045 CTGACCG 1104  
 Db 4587 GCGCTGCTGCG 4528  
 Oy 1105 GCGTTCGCGCGCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1164  
 Db 4527 GCGGATCTGTGCG 4475  
 Oy 1165 CCGCTGTGCG 1224  
 Db 4474 CGTGGAGCG 4420  
 Oy 1225 CCGCTGAGAGCG 1284  
 Db 4419 GCG 4360  
 Oy 1285 GGGGCGCGCGCGAGCTGAGCG 1344  
 Db 4359 GAGCCCG 4300

OY 1345 GTGGCCGATACCGCCACCTTCGACACCTTCACGAGCGCGCGCGG 1391  
 Db 4299 CCGG 4253

## RESULT 8

AAK53491

ID AAK53491 standard; DNA; 114955 BP.

AC AAK53491;

DT 05-JUL-1999 (first entry)

DE Human adenosine A1 receptor antisense oligonucleotide fragment.

KW Antisense oligonucleotide; multiple target; antisense treatment;

KW Impaired respiration; inflammation; lung disease;

KW Pulmonary vasoconstriction; inflammation; allergic rhinitis;

KW Acute asthma; allergy; asthma; impeded respiration;

KW Respiratory distress syndrome; pain; cystic fibrosis;

KW Pulmonary hypertension; pulmonary vasoconstriction; emphysema;

KW Chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;

KW Colon cancer; breast cancer; lung cancer; pancreatic cancer;

KW Hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;

KW Prostate cancer; ss.

KW Synthetic.

O5 MO9913886-A1.

PM 25-MAR-1999.

PD 17-SEP-1998; 98WO-US19419.

PR 09-JUN-1998; 98US-0093972.

PR 17-SEP-1997; 97US-0059160.

XX (TREC-) UNIV EAST CAROLINA.

PI NYCE JW;

DR WPI; 1999-229400/19.

XX New antisense oligonucleotides used in treatment of, e.g. pulmonary

XX vasoconstriction

XX Disclosure; Page 37; 120pp; English.

XX The specification describes antisense oligonucleotides (AAK52869-X55271)

XX directed against at least 2 mRNAs selected from target genes, coding and

XX non-coding regions of RNAs corresponding to target genes, gene

XX initiation codons, genomic flanking regions, intron-exon borders, the

XX 5'-end, the 3'-end and the juxta-section between coding and non-coding

XX or more diseases, conditions or mixtures. The antisense oligonucleotides

XX may be derived from sequences AAK5272-74. These multiple target

XX oligonucleotides (specifically AAK5272-74) can be used for the

XX antisense treatment of diseases and conditions. Typical diseases and

XX conditions are those associated with impaired respiration and

XX inflammation, including lung diseases, pulmonary vasoconstriction,

XX inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded

XX respiration, respiratory distress syndrome, pain, cystic fibrosis,

XX obstructive pulmonary disease (COPD), and cancers such as leukemias,

XX lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,

XX pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,

XX hepatic metastases, as well as all types of cancers which may metastasize

XX or have metastasized to the lungs, including breast and prostate cancer.

XX Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

XX Query Match 5.3%; Score 93.8; DB 20; Length 114955;

Best Local Similarity 32.6%; Pred. No. 1.2e-06;  
 Matches 500; Conservative 131; Mismatches 880; Indels 25; Gaps 5;

OY 44 CGATTCCAGACCCCTTCGACCTGCTGCTGCGGCGGACCTTCATGACGACGACAA 103  
 Db 104365 CCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104424  
 OY 104 CCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104424  
 Db 104425 GCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104424  
 OY 164 TGTGACGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104484  
 Db 104485 NCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104544  
 OY 224 TCATGACTGCGACACCGACGACGACGACGACGACGACGACGACGACGACGACG 283  
 Db 104545 NDNCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104604  
 OY 284 AGATCTC--GCAGGCGTCACGCGGTGTCACGCGGCAATGCGGCTACGCTGCGCGC 341  
 Db 104605 DNCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104664  
 OY 342 GCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 401  
 Db 104665 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104723  
 OY 402 TTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 461  
 Db 104724 GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104783  
 OY 462 GCGCTGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 521  
 Db 104784 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104843  
 OY 522 CGCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 581  
 Db 104844 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104903  
 OY 582 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 641  
 Db 104904 GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104963  
 OY 642 GATCATGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 701  
 Db 104964 NNDNCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105023  
 OY 702 CGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 761  
 Db 105024 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105073  
 OY 762 GACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 821  
 Db 105074 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105133  
 OY 822 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 881  
 Db 105134 NNDNCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105193  
 OY 882 GTATCCCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 941  
 Db 105194 NNDNCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105253  
 OY 942 CACCATATCATCTGTGTGCAAGCCCTTCCGGAATGAGCGGCGGCGGCGGCGGCGGCGG 1001  
 Db 105254 DNCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105313  
 OY 1002 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1061  
 Db 105314 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105373  
 OY 1062 CATCTACTTCATGATGAGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1121





OY 559 GCCGAGGAAGCCATGCGAGCGGCGCATCGGCAATTCGACCGGCGCTTCTACCGGCC 618  
 DB 24029 ACCGTCGGGCTGAGGAGTACCAACACCGCGCGCGCTGGCGCGGATACCGAA 24088  
 OY 619 GCCGCGCGCGCCACACCGAAGATCATCGAGGTGTCGCGCGCTGAGCGCGCATGCGC 678  
 DB 24089 GCCGCGCGCGCGACAGGCGCGCTTCAAGCGGTGTGCTCCGCTGCGCTGCGCACCGGC 24148  
 OY 679 GGCACTACAGGCAACCCATCGCGAGAGAGGAGACATCGTGGCGCGCGCTGAGGAA 738  
 DB 24149 GACGCGCGCGCACCGCGGTCGCGCGCGCGCTTCAACCTTCACACCGCGCGCTCCAGGCC 24208  
 OY 739 ACCTTCGCGCATGCGCGCGCGCGCATGAGTGGCGGCTGATCTCGACACCAAGATATG 798  
 DB 24209 CTCGCGCGAGCGCGCGCATGAGCGCGCGCTGTGGAACCTCACCGCGCGCGCTGTC 24268  
 OY 799 GGCGAGCGCAATTCGCGCGCGCTGCGCGAGACGCTGCGCGTGTATGAGCGCGCGCATGGCG 858  
 DB 24269 GGCGCGCGCGCAAGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 24319  
 OY 859 CGCGAGAGCGTCTCGCTGAGAGCGGTATCCCTAGTGGCGCGCGCTTCCACATGCTCAGAG 918  
 DB 24320 GGCGCGCGCGCTGCGCGCGCTGGAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24379  
 OY 919 GACCGCGTGTGCTGCGCGCGCGCGCATGATCATGCTGTCAGAGCGCGCTTCCCGGAACG 978  
 DB 24380 ACCCTGAGCG 24439  
 OY 979 AGCG 24496  
 DB 24440 GACG 24496  
 OY 1039 CCGGAGTGCAGCG 1098  
 DB 24497 GCCG 24556  
 OY 1099 ATCGTGGCGTGTGCG 1158  
 DB 24557 GCG 24616  
 OY 1159 CATCG 1215  
 DB 24617 CACTGCTGTGTACAGCG 24676  
 OY 1216 GCGCTTTCG 1275  
 DB 24677 GAACGTGAGGAATGCG 24736  
 OY 1276 GCGCTGCG 1335  
 DB 24737 GCGCTGCG 24796  
 OY 1336 CGGCG 1392  
 DB 24797 ACCG 24856  
 OY 1393 ATCGATTCGCTGTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1452  
 DB 24857 GTGCTGCG 24916  
 OY 1453 CATGCG 1512  
 DB 24917 CTCG 24973  
 OY 1513 GCGGTGAAGCG 1563  
 DB 24974 GGCACACTAGCG 25024

RESULT 10

AAL40781

ID AAL40781 standard; DNA: 88421 BP.

XX

AC AAL40781;  
 XX  
 DE 03-OCT-2002 (first entry)  
 XX  
 DE 88421nt genomic DNA of ramoplanin producing Actinoplanes sp.  
 XX  
 KW Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;  
 KW biosynthesis gene cluster; bioengineering; peptide synthetase module;  
 KW adenylation domain; hydroxyphenylglycine; Hpg; antibiotic precursor;  
 KW chlorinate; lipopeptide; gene; ds.  
 XX  
 OS Actinoplanes sp.  
 XX  
 FH Location/Qualifiers  
 FT 2077..3078  
 FT CDS  
 FT  
 FT /tag- a "Protein of ORF 1"  
 FT /product- 3118..4032  
 FT CDS  
 FT /tag- b "Protein of ORF 2"  
 FT /product- 4038..5048  
 FT CDS  
 FT /tag- c "Protein of ORF 3"  
 FT /product- complement (6665..5814)  
 FT CDS  
 FT /tag- d "Protein of ORF 4"  
 FT /product- complement (7703..6693)  
 FT CDS  
 FT /tag- e "Protein of ORF 5"  
 FT /product- complement (9464..8130)  
 FT CDS  
 FT /tag- f "Protein of ORF 6"  
 FT /product- 9691..10761  
 FT CDS  
 FT /tag- g "Protein of ORF 7"  
 FT /product- complement (12751..10829)  
 FT CDS  
 FT /tag- h "Protein of ORF 8"  
 FT /product- complement (13617..12802)  
 FT CDS  
 FT /tag- i "Protein of ORF 9"  
 FT /product- complement (15203..13614)  
 FT CDS  
 FT /tag- j "Protein of ORF 10"  
 FT /product- 15591..15863  
 FT CDS  
 FT /tag- k "Protein of ORF 11"  
 FT /product- 15880..19035  
 FT CDS  
 FT /tag- l "Protein of ORF 12"  
 FT /product- 19032..39713  
 FT CDS  
 FT /tag- m "Protein of ORF 13"  
 FT /product- 39713..65800  
 FT CDS  
 FT /tag- n "Protein of ORF 14"  
 FT /product- 65826..66530  
 FT CDS  
 FT /tag- o "Protein of ORF 15"  
 FT /product- 66546..67370  
 FT CDS  
 FT /tag- p "Protein of ORF 16"  
 FT /product- 67384..70059  
 FT CDS  
 FT /tag- q "Protein of ORF 17"  
 FT /product- 70099..70662  
 FT CDS  
 FT /tag- r "Protein of ORF 18"  
 FT /product- 70659..71906  
 FT CDS  
 FT /tag- s "Protein of ORF 19"  
 FT /product- complement (73439..71964)  
 FT CDS  
 FT /tag- t

FT CDS /product- "Protein of ORF 20"  
complement (74216..73563).  
/tag- u  
FT CDS /product- "Protein of ORF 21"  
complement (75424..74213)  
/tag- v  
FT CDS /product- "Protein of ORF 22"  
75535..76464  
/tag- w  
FT CDS /product- "Protein of ORF 23"  
complement (78110..76449)  
/tag- x  
FT CDS /product- "Protein of ORF 24"  
complement (79864..78107)  
/tag- y  
FT CDS /product- "Protein of ORF 25"  
complement (81624..79861)  
/tag- z  
FT CDS /product- "Protein of ORF 26"  
complement (81909..81682)  
/tag- aa  
FT CDS /product- "Protein of ORF 27"  
complement (82346..82062)  
/tag- ab  
FT CDS /product- "Protein of ORF 28"  
82587..84446  
/tag- ac  
FT CDS /product- "Protein of ORF 29"  
84481..85548  
/tag- ad  
FT CDS /product- "Protein of ORF 30"  
85556..86845  
/tag- ae  
FT CDS /product- "Protein of ORF 31"  
87372..86803  
/tag- af  
FT CDS /product- "Protein of ORF 32"  
87494..88420  
/tag- ag  
/product- "Protein of ORF 33"  
WO200231155-A2.  
PN 18-APR-2002.  
XX 15-OCT-2001; 2001WO-CA01462.  
XX 13-OCT-2000; 2000US-239924P.  
PR 12-APR-2001; 2001US-283296P.  
PR 24-JUL-2001; 2001US-0910813.  
XX (ECOP-) ECOPIN BIOSCIENCES INC.  
PA Farnet CM, Lazopoulos E, Staffa A;  
XX  
XX WPI: 2002-435445/46  
DR P-PSDB: AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,  
DR AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,  
DR AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,  
DR AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,  
DR AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.  
XX  
XX Novel isolated ramoplanin biosynthetic pathway polypeptide useful for  
PT chemically modifying biological molecule that is a substrate for a  
PT polypeptide encoded by a ramoplanin biosynthesis gene cluster -  
XX  
XX Disclosure; Page 87-135; 212pp; English.

CC molecule with the isolated polypeptide, where the polypeptide chemically  
CC modifies the biological molecule. The method comprises contacting the  
CC biological molecule with at least two different polypeptides encoded by  
CC ramoplanin ORFs 1-31. The polypeptides are useful for directing the  
CC biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated  
CC gene cluster comprising the ORFs is useful as a substrate for  
CC bioengineering of antibiotic structures. An isolated polypeptide or its  
CC encoding nucleic acid sequence is useful for generating derivatives of  
CC ramoplanin, for improving production or for producing variants of other  
CC antibiotics of the peptide class. The isolated polypeptides are useful  
CC for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain  
CC in conjunction with other peptide synthetase modules and allowing the  
CC incorporation of Thr into a peptide antibiotic precursor, for modifying  
CC fatty acid structure and/or enhancing fatty acid incorporation into the  
CC peptide antibiotic structure, for production of an hydroxyphenylglycine  
CC (HPG)-containing peptide antibiotic, for enhancing secretion of  
CC ramoplanin or its variants and derivatives, for enhancing uptake of  
CC precursors for ramoplanin biosynthesis, for enhancing production of  
CC ramoplanin products or its variants or derivatives, to chlorinate HPG of  
CC a peptide antibiotic precursor, and for designing specific nucleotide  
CC probes and primers for identifying and isolating putative lipopeptide  
CC -producing microorganisms. This polynucleotide sequence represents the  
CC 88421nt genomic DNA of a ramoplanin producing Actinoplanes sp.  
CC  
XX

SO Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 other;

Query Match 4.9%; Score 86; DB 24; Length 88421;  
Best Local Similarity 43.1%; Pred. No. 2e-05; Mismatches 775; Indels 24; Gaps 3;

Matches 605; Conservative 0; Mismatches 775; Indels 24; Gaps 3;

QY 35 TGTCCCAATCCGATTCACACCCCTTGTGCTGCTGCGCGGCGACCTCATGACG 94  
DB 46230 TGGCCCGGCGAGGTGCTGCGACAGGTGTCACGGTGCAGCGGCGACCGTGCAC 46239  
QY 95 GCAGCAACACCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154  
DB 46290 ACATCTCCGACCCCG 46349  
QY 155 TGGCGATCTGTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 214  
DB 46350 GGGTCAGAGGCGGCG 46409  
QY 215 CGCGCGCTTCAATGACATGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 268  
DB 46410 CGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 46469  
QY 269 GCGACATGACGCCAAGATCTGCGAGGCGGTACACGCGTGTACAGGCAATTCGGCA 328  
DB 46470 ACCAATCG 46529  
QY 329 TCAGCTGCG 388  
DB 46530 ACG 46589  
QY 389 GCGGCTTTCACGCTTTCAGAGCGCTTTCGCGCACTACGAGCGGTGCGCGCGCGCG 448  
DB 46590 ACTATGCGCGCTTTCGAGGCGGTACCTGCGCTGCGAGCGAGACCGCGCGCGCGCT 46649  
QY 449 CGGCGCTCAAGCGCGCGCTGTATGAGGCGCATTCAGCGTGCAGGCGCGCGCGCT 508  
DB 46650 CCGCGCAATGCG 46709  
QY 509 ACTTGCAGCG 568  
DB 46710 CGCGCGACCG 46769  
QY 569 CCATGCGAGCG 628  
DB 46770 TCCG 46829  
QY 629 CCACACCGAGAGATCATGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 688

FT	RBS	/product- "EvdA"
FT	complement (1389..1394)	
FT	tag- b	
FT	complement (1490..2611)	
FT	tag- c	
FT	/product- "EvdB"	
FT	complement (2618..2622)	
FT	tag- d	
FT	complement (2622..3860)	
FT	tag- e	
FT	/product- "EvdC"	
FT	complement (3867..3870)	
FT	tag- f	
FT	4143..5312	
FT	tag- g	
FT	/product- "EvdD"	
FT	4134..4138	
FT	tag- h	
FT	5309..6235	
FT	tag- i	
FT	/product- "EvdE"	
FT	6332..7275	
FT	tag- j	
FT	/product- "EvdF"	
FT	6226..6229	
FT	tag- k	
FT	7272..8327	
FT	tag- l	
FT	/product- "EvdG"	
FT	8342..9364	
FT	tag- m	
FT	/product- "EvdH"	
FT	8333..8336	
FT	tag- n	
FT	complement (9463..10224)	
FT	tag- o	
FT	/product- "EvdI"	
FT	complement (10232..10235)	
FT	tag- p	
FT	10424..11176	
FT	tag- q	
FT	/product- "EvdJ"	
FT	12027..12455	
FT	tag- r	
FT	/product- "EvdK"	
FT	partial	
FT	/note- "No start codon"	
FT	complement (12108..13022)	
FT	tag- s	
FT	/product- "EvdL"	
FT	complement (13027..13030)	
FT	tag- t	
FT	complement (14410..15363)	
FT	tag- u	
FT	/product- "EvdA"	
FT	complement (15369..15373)	
FT	tag- v	
FT	complement (15380..16414)	
FT	tag- w	
FT	/product- "EvdB"	
FT	complement 16415..17873	
FT	tag- x	
FT	/product- "EvdC"	
FT	complement (17870..18934)	
FT	tag- y	
FT	/product- "EvdD"	
FT	19374..20906	
FT	tag- z	
FT	/product- "EvdE"	
FT	21064..22542	
FT	tag- aa	
FT	/product- "EvdF"	
FT	21056..22542	
FT	tag- ab	
FT	/product- "EvdG"	
FT	21056..22542	
FT	tag- ac	
FT	/product- "EvdH"	
FT	21056..22542	
FT	tag- ad	
FT	/product- "EvdI"	
FT	21056..22542	
FT	tag- ae	
FT	/product- "EvdJ"	
FT	21056..22542	
FT	tag- af	
FT	/product- "EvdK"	
FT	21056..22542	
FT	tag- ag	
FT	/product- "EvdL"	
FT	21056..22542	
FT	tag- ah	
FT	/product- "EvdM"	
FT	21056..22542	
FT	tag- ai	
FT	/product- "EvdN"	
FT	21056..22542	
FT	tag- aj	
FT	/product- "EvdO"	
FT	21056..22542	
FT	tag- ak	
FT	/product- "EvdP"	
FT	21056..22542	
FT	tag- al	
FT	/product- "EvdQ"	
FT	21056..22542	
FT	tag- am	
FT	/product- "EvdR"	
FT	21056..22542	
FT	tag- an	
FT	/product- "EvdS"	
FT	21056..22542	
FT	tag- ao	
FT	/product- "EvdT"	
FT	21056..22542	
FT	tag- ap	
FT	/product- "EvdU"	
FT	21056..22542	
FT	tag- aq	
FT	/product- "EvdV"	
FT	21056..22542	
FT	tag- ar	
FT	/product- "EvdW"	
FT	21056..22542	
FT	tag- as	
FT	/product- "EvdX"	
FT	21056..22542	
FT	tag- at	
FT	/product- "EvdY"	
FT	21056..22542	
FT	tag- au	
FT	/product- "EvdZ"	
FT	21056..22542	
FT	tag- av	
FT	/product- "EvdAA"	
FT	21056..22542	
FT	tag- aw	
FT	/product- "EvdAB"	
FT	21056..22542	
FT	tag- ax	
FT	/product- "EvdAC"	
FT	21056..22542	
FT	tag- ay	
FT	/product- "EvdAD"	
FT	21056..22542	
FT	tag- az	
FT	/product- "EvdAE"	
FT	21056..22542	
FT	tag- ba	
FT	/product- "EvdAF"	
FT	21056..22542	
FT	tag- bb	
FT	/product- "EvdAG"	
FT	21056..22542	
FT	tag- bc	
FT	/product- "EvdAH"	
FT	21056..22542	
FT	tag- bd	
FT	/product- "EvdAI"	

```

/Product- "EvDA"
  complement (1389..1394)
  /tag- b
    complement (1490..2611)
  /tag- c
    /product- "EvDB"
      complement (2618..2622)
  /tag- d
    complement (2622..3860)
  /tag- e
    /product- "EvDC"
      complement (3867..3870)
  /tag- f
    4143..5312
  /tag- g
    /product- "EvDD"
      4134..4138
  /tag- h
    5309..6335
  /tag- i
    /product- "EvDE"
      6232..7275
  /tag- j
    /product- "EvDF"
      6226..6229
  /tag- k
    7272..8327
  /tag- l
    /product- "EvDG"
      8342..9364
  /tag- m
    /product- "EvDH"
      8333..8336
  /tag- n
    complement (9463..10224)
  /tag- o
    /product- "EvDI"
      complement (10232..10235)
  /tag- p
    10424..11176
  /tag- q
    /product- "EvDJ"
      12027..12455
  /tag- r
    /product- "EvDK"
      partial
  /note- "No start codon"
    complement (12108..13022)
  /tag- s
    /product- "EvDL"
      complement (13027..13030)
  /tag- t
    complement (14410..15363)
  /tag- u
    /product- "EvEA"
      complement (15369..15373)
  /tag- v
    complement (15380..16414)
  /tag- w
    /product- "EvEB"
      complement 16419..17873
  /tag- x
    /product- "EvEC"
      complement (17870..18934)
  /tag- y
    /product- "EvED"
      19374..20906
  /tag- z
    /product- "EvFE"
      21064..22542
  /tag- aa
    /product- "EvFF"
      21056..22542

```

```

FT      CDS      /tag- ab
FT      22748..24172
FT      /tag- ac
FT      /product- "EvrG"
FT      22736..22740
FT      /tag- ad
FT      complement (24177..25223)
FT      /tag- ae
FT      /product- "EvrH"
FT      complement (25230..25233)
FT      /tag- af
FT      25350..26626
FT      /tag- ag
FT      /product- "EvrI"
FT      26685..30479
FT      /tag- ah
FT      /product- "EvrJ"
FT      26672..26676
FT      /tag- ai
FT      complement (30557..31876)
FT      /tag- aj
FT      /product- "EvrK"
FT      complement (31885..31888)
FT      /tag- ak
FT      complement (31941..32882)
FT      /tag- al
FT      /product- "EvrL"
FT      complement (33167..34405)
FT      /tag- am
FT      /product- "EvrM"
FT      complement (34414..34418)
FT      /tag- an
FT      complement (34449..35210)
FT      /tag- ao
FT      /product- "EvrN"
FT      complement (35219..35221)
FT      /tag- ap
FT      complement (35294..36238)
FT      /tag- aq
FT      /product- "EvrO"
FT      complement (36235..36963)
FT      /tag- ar
FT      /product- "EvrP"
FT      complement (36998..38026)
FT      /tag- as
FT      /product- "EvrQ"
FT      complement (38072..38566)
FT      /tag- at
FT      /product- "EvrR"
FT      complement (38992..40163)
FT      /tag- au
FT      /product- "EvrS"
FT      complement (40216..40890)
FT      /tag- av
FT      /product- "EvrT"
FT      complement (40899..40902)
FT      /tag- aw
FT      complement (40887..41576)
FT      /tag- ax
FT      /product- "EvrU"
FT      complement (41679..42707)
FT      /tag- ay
FT      /product- "EvrV"
FT      complement (42714..42717)
FT      /tag- az
FT      complement (42810..43799)
FT      /tag- ba
FT      /product- "EvrW"
FT      complement (43807..43811)
FT      /tag- bb
FT      complement (43799..44866)
FT      /tag- bc
FT      /product- "EvrX"

FT      CDS      complement (45014..45760)
FT      /tag- bd
FT      /product- "EvrY"
FT      complement (45767..45770)
FT      /tag- be
FT      complement (45962..46714)
FT      /tag- bf
FT      /product- "EvrZ"
FT      complement (45952..45956)
FT      /tag- bg
FT      complement (47156..49234)
FT      /tag- bh
FT      /product- "EvrA"
FT      51627..52715
FT      /tag- bi
FT      /product- "EvrB"
FT      51629..51622
FT      /tag- bj
FT      52889..53557
FT      /tag- bk
FT      /product- "EvrC"
FT      53554..54207
FT      /tag- bl
FT      /product- "EvrD"
FT      complement (54362..55117)
FT      /tag- bm
FT      /product- "EvrE"
FT      complement (55125..55128)
FT      /tag- bn
FT      complement (55135..56094)
FT      /tag- bo
FT      /product- "EvrF"
FT      complement (56100..56103)
FT      /tag- bp
FT      complement (56184..56813)
FT      /tag- bq
FT      /product- "EvrG2"
FT      56961..58709

Query Match      4.8%; Score 84.6; DB 22; Length 109519;
Best Local Similarity 45.1%; Pred. No. 3.2e-05;
Matches 542; Conservative 0; Mismatches 629; Indels 30; Gaps 5;

QY      117 CGCGACCTGGGCGTGGCGGCGACCGCATCGCCCATCGCGCATCTGTGAGCGCGC 176
DB      94498 CGTGGACCTTGTTCACAGGCGCGCGACCTGCGCGTTCGTCACCGGTACGCGCT 94557
QY      177 CGCGCACACCGCGGTGACGTGTGGGCTGTGTGTCGCGCGCGCTTCATCGACTCGCA 236
DB      94558 GGAACCGCGCGCGATCGGCGCGCGCGCTGCGCGGTGGAACGCGCTCGCGCAC 94617
QY      237 CACCCACGACGACACATCTACCTCTAGAGCGTGGCGACATGACGCCCAAGACTGCAAGG 296
DB      94618 CGGCTTCGCGGTGAGACCGCGCGCGCGCGCTGTCTCGCGGTGACCGCGCACGCTGT 94677
QY      297 CGTCAACAGGTGTGTACAGGCGCAATTGCGGCATCAGCTGCGCGCGTGGCGACGCCAA 356
DB      94678 GGTCAACCGCGTGGCGGACGCGGCGTGTGCGGGGTGTGTGTCGGGTCTCGGGGTTC 94737
QY      357 CCGCGCGCGCGCGCGTGGACCTGTGTGACGAAGCGCGCTTACCGTT-----T 404
DB      94738 CTTGGCGCGCGCTGAGGTGGCGTGGCGCTGTGTGCGGCGCGCGGACGACCCAGCT 94797
QY      405 CGAGCGCTTCGCGCATCTGACGCGGTGGCGGCGCACCGCGCGCGCGCTCAACGCCGC 464
DB      94798 CCGGCGCGCGGTGTGTGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 94857
QY      465 CTGTATGTGTGGCCATTCACGCTGCGCGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 524
DB      94858 CGCACCGCGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 94917
QY      525 CACGACGAGGAATTCGCGCGCATGCGCGGACCTGTGCGGAGAACCATGTGCGCACGCGCG 584

```

FT	Key	Location/Qualifiers
FT	CDS	1..1248
FT		/'tag'-a
FT		/product="Desi"
PN		MO200000620-A2.
PD		06-JAN-2000.
XX		
PF	25-JUN-1999;	99MO-US14398.
XX		
PR	26-JUN-1998;	9805-0105537.
XX		
PA	(MINU )	UNIV MINNESOTA.
PI		
PI	Sherman DH, Liu H, Xue Y, Zhao L;	
XX		
DR	WPI, 2000-160679/14.	
P-PSDB;	AA77181.	
PT	Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.	
XX	synthesis of methymycin and pikromycin	
PS	Claim 3; Page 353-354; 438pp; English.	
XX		
CC	The invention relates to an isolated and purified nucleic acid segment	
CC	comprising a desosamine biosynthetic gene cluster, a fragment or its	
CC	biologically active variant, where the nucleic acid sequence is not	
CC	derived from the eryc gene cluster of <i>Saccharopolyspora erythraea</i> or	
CC	<i>Streptomyces antibioticus</i> . The invention also relates to a macrolide	
CC	biosynthetic gene cluster, or fragments thereof. The macrolide	
CC	biosynthetic gene cluster encodes proteins which synthesize methymycin,	
CC	pikromycin, neomethymycin, nadomycin or a combination of these	
CC	compounds. Recombinant or augmented cells comprising the desosamine	
CC	and/or macrolide biosynthetic gene clusters are useful for the	
CC	production of biologically active macroclides. The macrolide biosynthetic	
CC	proteins are useful for synthesis of methymycin, pikromycin,	
CC	neomethymycin and nadomycin. The alternative termination of polyketide	
CC	synthesis may be useful to prepare novel antibiotics and	
CC	polyhydroxyalanate (PHA) monomers. The compounds produced by the	
CC	recombinant host cells are useful as biopolymers, e.g., in packaging or	
CC	biomedical applications, to engineer PHA monomer syntheses or to prepare	
CC	biologically active agents, such as chemotherapeutics,	
CC	immunosuppressants, agents to treat asthma, chronic obstructive pulmonary	
CC	disease as well as other diseases involving respiratory inflammation,	
CC	cholesterol-lowering agents or macroclide-based antibiotics which are	
CC	active against a variety of organisms, e.g., bacteria, including	
CC	multi-drug resistant pneumococci and other respiratory pathogens, as well	
CC	as viral parasitic pathogens, or as crop protection agents (e.g.,	
CC	fungicides or insecticides) via expression of polyketides in plants.	
CC	Sequences AA87286-287224 represent desosamine biosynthetic genes from	
CC	<i>Streptomyces venezuelae</i> ATCC 15439, which encode proteins	
CC	AA77181-777189.	
XX		
XX		
SO	Sequence 1248 BP; 177 A; 546 C; 377 G; 148 T; 0 other;	
Query Match	4.7%; Score 81.8; DB 21; Length 1248;	
Best Local Similarity	44.2%; Pred. No. 0.00015;	
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;		
0Y	32 CCAATGCCAATCCGATTCGCCAGCCCTTCGACTCTGCTCCGGGCGGACCCATCATCG	91
Db	215 CCACCTGCACAGCCACGCGCGGGCTTCAGACTCTCGCGCAGCCCGCCCTCACCGGCG	274
0Y	92 ACGGAGCAACACCCCGGGCGCGCGCGCGGACCTGGGCTGGCGGCGACCGCATCGCGG	151
Db	275 AAGTGATATGCGCGCTGCATGACGTTTCGCCGCGCACCCCGACGCACTGCGGTGATCGGCG	334
0Y	152 CCATGCGGATCTGTGAGACCGCGCGCGACACCCGGGTGCAGCTGTGCGGCGGTGG	211
Db	335 TCACCGCGGTCTTGCGCGACATCGACCGCGACACCGGACACCTGCAGCCCGGACGAGTGG	394
0Y	212 TCGCGCGCGGCTTATGACTGCACTGCACACCCAGACGACGACACTGCTGCTGAGGCTCGG	271

Db 395 CCGCGCGGTCACACCCCGACCTCGCGCGTGTGCGGCTGCACCTCTGGGGCCGCGCTT  
 QY 272 ACATAGCGCCCAAGATCTGCGAGGGCGTTCACACAGTGTGTCACGGCAATTGCGGCATCA 331  
 Db 455 GCGCGCGGACCAAGCTGCGGAAAGTGGCGAGACAGACGGCTGCGGCTGTACTTTCAGACG 514  
 QY 332 GCGTGGCGCGGCTGGCGGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391  
 Db 515 CCGCGCAGCGCCCTGCGGTCGCGGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 574  
 QY 392 GCTCTTACCGTTTGCAGCGCTTCCCGCACTACCTGAGCGGCTTGGCGGCGCACCGCGCG 451  
 Db 575 AGGTCTTACGCTTCA--CCGCAACCAAGCGCGTCAACGCCCTTCAGAGGGCGCGCGCTG 631  
 QY 452 CCGTCAACGCGCGCGTGTGAGTGGGCGCATTCACCGCTCGCGCGCGCGCGCGCGCGCGCT 511  
 Db 632 TCACCGAGAGAGCGGACCTCCCGCGCGGATCCCGCGCGCGCGCGCGCGCGCGCGCGCG 691  
 QY 512 TGCAGCG 571  
 Db 692 TCG 751  
 QY 572 TGGCGAGCG 631  
 Db 752 TGGCGCTACCTCTCTGACCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 811  
 QY 632 CCACCGAAGATCATGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691  
 Db 812 CCGCGCTACCGCGAGACCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 871  
 QY 692 CCCACATGCGCGAGAGAGCGCGAGCGACATCTGTGCGCGCGCGCGCGCGCGCGCGCG 751  
 Db 872 ACGGCGCAACACACACACATACGTATCGTACGATGACGAGAGCGACACCGCGCGCG 931  
 QY 752 GCGCGAGCTGAGAGCG 811  
 Db 932 ACGCGAGCTGCTGATGAGAGTCTTGAAGCGCGAGCGCGCGCGCGCGCGCGCGCGCG 991  
 QY 812 TCGCGCGCTGCGCGAGACCTGCGCGCTGATTCGAG-----GCGCGCATGCGCGCGCG 862  
 Db 992 CCGCGCGCTGCGCGAGCG 1051  
 QY 863 AGGACGCTGCTGCGAGCG 922  
 Db 1052 CCGAGCGCTGCG 1111  
 QY 923 GCGTGTGCTGCGCGAGCG 982  
 Db 1112 ACATCG 1171  
 QY 983 GCGCGAGCTGATGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015  
 Db 1172 CCGCGCAGCG 1204

## RESULT 13

AAD39044  
 ID AAD39044 standard; DNA; 1248 BP.

AAD39044;

23-SEP-2002 (first entry)

DE Streptomyces venezuelae Desi gene.

KW Glycosylated polyketide; modified recombinant bacterial host cell;

KW mRHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;

KW tetracycline; polyene; polyether; ansamycin; isochromanquinone; sugar;

KW deosamine; Desi; 4-dehydrase; enzyme; gene; ds.

OS Streptomyces venezuelae.

FH Key Location/Qualifiers  
 FT CDS 1..1248  
 FT /tag= a  
 FT /product= "Streptomyces venezuelae Desi protein"  
 FT /trans\_except= (pos:1..3, aa:Met)  
 FT /note= "CDS does not include start codon"  
 FT /partial  
 W0200229035-A2.  
 11-APR-2002.  
 05-OCT-2001; 2001WO-US31255.  
 05-OCT-2000; 2000US-238185P.  
 (MINU ) UNIV MINNESOTA.  
 PA (LUDH.) LIU H.  
 PA (SHER.) SHERMAN D H.  
 PA (ZHAO.) ZHAO L.  
 PI Liu H, Sherman DH, Zhao L;  
 DR WPI: 2002-405171/43.  
 DR P-PSDB: AAE24229.  
 XX  
 PT Modified recombinant bacterial host cells in which the expression and  
 PT activity of nucleic acids encoding sugar biosynthetic enzymes has been  
 PT altered, useful for producing metabolites with altered sugar structures  
 PT  
 PS Disclosure; Page 153-154; 174pp; English.  
 CC The invention provides a method to alter the sugar structure diversity  
 CC for a particular metabolite via the recruitment and collaborative action  
 CC of sugar genes from a variety of sugar biosynthetic pathways to yield a  
 CC metabolite comprising a non-natural sugar, e.g., a novel glycosylated  
 CC polyketide. The invention also relates to a modified recombinant  
 CC bacterial host cell (mRHC) in which the expression and activity of  
 CC nucleic acids encoding sugar biosynthetic enzymes has been altered.  
 CC The mRHCs may be cultured to produce the modified sugar products,  
 CC e.g., a macrolide, anthracycline, angucycline, avermectin, milbemycin,  
 CC tetracycline, polyene, polyether, ansamycin or isochromanquinone.  
 CC The present sequence is Streptomyces venezuelae sugar (deosamine)  
 CC biosynthetic gene cluster Desi (4-dehydrase) gene.  
 XX  
 SQ Sequence 1248 BP; 177 A; 546 C; 377 G; 148 T; 0 other;  
 Query Match 4.7%; Score 81.8; DB 24; Length 1248;  
 Best Local Similarity 44.2%; Pred. No. 0.00015;  
 Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;  
 QY 32 CCATGTCCTCCATTCGATCCCAAGCCCTTGCAGCTGCTGCGGCGCGCGCGCGCGCGCG 91  
 Db 215 CCACGTGCAACGCGACG 274  
 QY 92 ACGGCAACACCCCG 151  
 Db 275 AAGTATCATGCGCGTGTGATGAGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 334  
 QY 152 CCATGCGGATCTGTGCG 211  
 Db 335 TCACCGCGGCTTCCG 394  
 QY 212 TCGCGCGCGCGCTTCACTGCACTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 271  
 Db 395 CCGCGCGGCTACACCG 454  
 QY 272 ACATAGCGCCCAAGATCTGCGAGGGCGTTCACAGGCTGTACAGGCAATTGCGGCATCA 331  
 Db 455 GCGCGCGGACCAAGCTGCGGAGGTGCGCGAGACAGCGCGCTGCGGTGATTCGACG 514  
 QY 332 GCGTGGCGCGGCTGGCGGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391



[illegible]

OY 452 CCGTCAAGCGCCGCTGATGGGCGCATTCACGCTGCGCGCCGCTATGCCGACT 511  
 Db 11565 TCACCGAGAGACCGGACCTCCCGCCGAGTCCGCGCCCTCCACAACTTCGGCTTGAC 11506  
 OY 512 TCGACGCGCGCCGCGACGAGGAAATCGCGCCATCGGAGCTGGCGAGAAAGCA 571  
 Db 11505 TCGCGGCGGAGCGCCGCGCGGAGCAAGCCAAATAGAGAGAGCCCGCCGCCA 11446  
 OY 572 TGGCGACGCGCGCATGGCGATTTCGACCGGCGCTTACCGCCCGCCCGCGCA 631  
 Db 11445 TGGGCGCTACCTCCCTCGACCGTTCCTCCGAGTCTATCGAACCGGCGCAACAG 11386  
 OY 632 CCACCGAAGATCATGAGAGTGTGCGCGCGCGCTGAGCGCCATGCGCGCATCTAC 691  
 Db 11385 CCGCTTACCGCGAGACCTCGCGACCTCCCGCGCTCTCTGCGCGCACGACGCGC 11326  
 OY 692 CCCACATCGCGAGAGGAGCGACATCTGCGCGCGCTGAGGAAACCTTCGCGATCG 751  
 Db 11325 ACGGCCCTCAACACACAGAGTGTGAGTGTGAGAGAGAGCGCACCGCGCATTC 11286  
 OY 752 GCGCGGAGTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 811  
 Db 11265 ACCGCGACCTGTGTATGAGAGTGTGAGAGCGGAGCGTCAACACCGCGCTACT 11206  
 OY 812 TCGGCGCGCTGCGCGAGAGCGCTGCGCTGATCGAG-----GCCGCGCATGGCGCGC 862  
 Db 11205 CCGCGGCTGTCCACAGAGCTGAGAGCGGAGCGCGCGCGCGCGCGCGCGCA 11146  
 OY 863 AGACGCTGTGTGTGAGAGCGGTATCCCTACGTGCGCGCGCTCCACCATGCTCAAGAG 922  
 Db 11145 CCGAGACGCTGTGCGCGCGCGGTGTGCTCCGCGCGAGCGGAGCGCATCGCGAG 11086  
 OY 923 GCGTGTGTGTGCGCGCGCGAGCGACATCATCTGTGTGTGTGTGTGTGTGTGTGT 982  
 Db 11085 ACATCGCGCGGT 11026  
 OY 983 GCGCGACGCTGT 1015  
 Db 11025 CCGCGCACCGCGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10993

RESULT 15  
 AA287319/c  
 ID AA287319 standard; DNM; 13613 BP.  
 XX  
 AC AA287319;  
 XX  
 DT 05-JUN-2000 (first entry)  
 XX  
 DE  
 XX  
 S. venezuelae desosamine biosynthetic gene cluster plkB.  
 KM Desosamine biosynthesis; macroide; polyketide; methylmycin; pikromycin;  
 KM neomethylmycin; narbomycin; polyhydroxyalkanoate monomer synthesis;  
 KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;  
 KM chronic obstructive pulmonary disease; respiratory inflammation;  
 KM hypercholesterolemia; crop protection agent; ds.  
 XX  
 OS Streptomyces venezuelae ATCC15439.  
 XX  
 FH Key  
 FH CDS Location/Qualifiers  
 FT 1..809  
 FT /tag- a  
 FT /product- "plkB gene cluster protein #1 (AAV77204)"  
 FT /note- "No initiation codon given in the specification"  
 FT 806..2014  
 FT /tag- b  
 FT /product- "plkB gene cluster protein #2 (AAV80998)"  
 FT 2162..6741  
 FT /tag- c  
 FT /product- "plkB gene cluster protein #3 (AAV77205)"  
 FT 6834..7402  
 FT CDS

FT /tag- d  
 FT /product- "plkB gene cluster protein #4 (AAV77206)"  
 FT /transl\_except- (pos:6837..6841, aa:Gln)  
 FT 7492..8205  
 FT CDS  
 FT /tag- e  
 FT /product- "plkB gene cluster protein #5 (AAV77207)"  
 FT complement (7942..8205)  
 FT /partial  
 FT /tag- f  
 FT /product- "plkB gene cluster protein #6 (AAV77208)"  
 FT /note- "No termination codon given in the specification"  
 FT /transl\_except- (pos:8270..8272, aa:Val)  
 FT /transl\_except- (pos:8273..8275, aa:Thr)  
 FT /transl\_except- (pos:8276..8278, aa:Gly)  
 FT complement (10126..11139)  
 FT CDS  
 FT /tag- g  
 FT /product- "plkB gene cluster protein #7 (AAV80999)"  
 FT complement (11271..12149)  
 FT /tag- h  
 FT /product- "plkB gene cluster protein #8 (AAV77209)"  
 FT complement (12342..13799)  
 FT CDS  
 FT /tag- i  
 FT /product- "plkB gene cluster protein #9 (AAV77210)"  
 FT complement (13706..15043)  
 FT CDS  
 FT /tag- j  
 FT /product- "plkB gene cluster protein #10 (AAV77211)"  
 FT complement (15404..15574)  
 FT /tag- k  
 FT /product- "plkB gene cluster protein #11 (AAV77212)"  
 FT  
 FT W0200000620..A2.  
 FT  
 FT 06-JAN-2000.  
 FT  
 FT 25-JUN-1999; 99WO-0514398.  
 FT  
 FT 26-JUN-1998; 98US-0105337.  
 FT  
 FT (MIND ) UNIV MINNESOTA.  
 FT  
 FT Sherman DH, Liu H, Xue Y, Zhao L;  
 FT  
 FT WPI: 2000-160679/14.  
 FT  
 FT P-PSDB; AAV77204, AAV77205, AAV77206, AAV77207, AAV77208,  
 FT AAV77209, AAV77210, AAV77211, AAV77212, AAV80998, AAV80999.  
 FT  
 FT Desosamine and macroide biosynthetic gene clusters, useful for, e.g.  
 FT synthesis of methylmycin and pikromycin.  
 FT  
 FT Disclosure; Figure 32; 438pp; English.  
 FT  
 FT The invention relates to an isolated and purified nucleic acid segment  
 FT comprising a desosamine biosynthetic gene cluster, a fragment or its  
 FT biologically active variant, where the nucleic acid sequence is not  
 FT derived from the eyc gene cluster of *Saccharopolyspora erythraea* or  
 FT *Streptomyces antibioticus*. The invention also relates to a macroide  
 FT biosynthetic gene cluster, or fragments thereof. The macroide  
 FT biosynthetic gene cluster encodes proteins which synthesise methylmycin,  
 FT pikromycin, neomethylmycin, narbomycin or a combination of these  
 FT compounds. Recombinant or augmented cells comprising the desosamine  
 FT and/or macroide biosynthetic gene clusters are useful for the production  
 FT of biologically active macroides. The macroide biosynthetic proteins  
 FT are useful for synthesis of methylmycin, pikromycin, neomethylmycin and  
 FT narbomycin. The alternative termination of polyketide synthesis may be  
 FT useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)  
 FT monomers. The compounds produced by the recombinant host cells are useful  
 FT as biopolymers, e.g., in packaging or biomedical applications, to  
 FT engineer PHA monomer synthases or to prepare biologically active agents,  
 FT such as chemotherapeutics, immunosuppressants, agents to treat asthma,  
 FT chronic obstructive pulmonary disease as well as other diseases involving  
 FT respiratory inflammation, cholesterol-lowering agents or macroide-based  
 FT antibiotics which are active against a variety of organisms, e.g.,  
 FT bacteria, including multi-drug resistant pneumococci and other

CC respiratory pathogens, as well as viral parasitic pathogens, or as crop  
 CC protection agents (e.g., fungicides or insecticides) via expression of  
 CC polyketides in plants. The present sequence represents the desamine  
 CC biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC  
 CC 15439, as given in figure 32.  
 XX

Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other;

## Query Match

Best Local Similarity 4.7%; Score 81.8; DB 21; Length 13613;  
 Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

QY 32 CCATGTCCTCCATCCGATCCGAGCCCTTGACCTGCTGCGGGGCGGACCTTCATCG 91  
 DB 12669 CCACCTGGCAACGCGCAGCGCGGCTTCACCTCTGCGCGACCGCGGCTCACCGGCG 12610  
 QY 92 ACGGCAAGCAACCCCGGGGCGCGCGACCTGGGCGTGGCGGCGGACCGCATCGCGG 151  
 DB 12609 AAGTATCATGCGCTGTGATGACGTTCGCGCGCACCCCGACGCACTGCGCTGATCGGCG 12550  
 QY 152 CCATGCGGCGATCTGTGCGAGCGCGCGCGGCGACACCGGCGTGTGCGGCTGTGG 211  
 DB 12549 TCACCGCGGCTTCGCGCGACATGACCGGACCGGACCGGACCTCGACCGGACGAGTGG 12490  
 QY 212 TCGCGCGCGGCTTCATGCTGCTGACACCGACGAGACACTGCTGCTCAAGCGTGGCG 271  
 DB 12489 CCGCGCGGCTGACACCGCGCACCTGCGCGCTGTGCGCTGCTGCGCGCGCGCTT 12430  
 QY 272 ACATGACGCCCAAGATCTGCGAGGCGGTACACGAGGTGTACGCGGCAATTGCGGATCA 331  
 DB 12429 GCGCGCGCGACCTGCGGAGGCTGCGCGAGGACGAGGACGCGCTGCGGTGTACTTCGACG 12370  
 QY 332 GCGTGGCGCGCTGCGCGGCGCAACCCCGCGCGCGCGCGCTGACCTGCTGAGAGGCG 391  
 DB 12369 CCGGCGACGCGCTGCGCGCTGCGCGGTGCGAGCGCGCGCGCGCGCTTGGCGACGCG 12310  
 QY 392 GCTTTACCGTTTGGAGCGCTTCGCGACTGACGCGGTGCGGGGCGGCGCGCGCGG 451  
 DB 12309 AGGTCTTCAGCTTCCA---CGCCACCAAGGCGCTTCACACGCTTGAAGGCGCGCGTGC 12253  
 QY 452 CCGTCAACGCGCGCTGTATGTTGGGCGATTCAACGCTGCGCGCGCGCGCTGATGCGGACT 511  
 DB 12252 TCACCGGAGAGCGCGGACCTGCGCGCGCGGATCGCGCGCTTCACACTTGGGTTGACG 12193  
 QY 512 TGCAGCGCGCGCGCGCGGAGAGAAATGCGGCGCATGCGGAGCTGCGGAGAGAGCA 571  
 DB 12192 TGC CGCGGCGGAGCGCGCGCGGCGGAGCAACGCGCAAGATGAGCGCGCGCGCCA 12133  
 QY 572 TGGCGAGCGCGCGCATGCGGATTCGACGCGCGCTTCTACCGCGCGCGCGCGCGCGCA 631  
 DB 12132 TGGGCGCTCACTCTCTGAGCGCTTTCGCGAGGTCAITCGACCGGAACCGCGCACGCG 12073  
 QY 632 CCACGAGAGATCATGAGGTGTGCGGCGCGGTGAGCGGCGGCGCATCTACGCA 691  
 DB 12072 CCGCGCTACGCGGAGCACTGCGGAGCTCCCGCGCTCTGTCGCGGACGAGCGCGCG 12013  
 QY 692 CCACATGCGCGAGAGGCGAGACATGTGCGCGCGCGGTGAGAGAACTTCCGATCG 751  
 DB 12012 ACGGCGCTCAACAACGAGTGTGATGTGAGATGACGAGGCGCACCGCGGATCC 11953  
 QY 752 GCGCGAGAGTGAAGTGTGCGGTGTGATCTGACACCAAGGTCAATGCGCGAGCGCAATT 811  
 DB 11952 ACCGCGACTCTGTATGAGGTCTGAGAGCGGAGGCGGTGACACCGCGCGCTACTTCT 11893  
 QY 812 TCGGCGGCTGCGCGGAGAGCGCGCGGTGATGAG-----GCGCGCATGCGCGCG 862  
 DB 11892 GCGCGGCGTCCGACAGCTGAGCGGTGACCGCGGCGAGCGCGCGCGCTGCGGACA 11833  
 QY 863 AGGAGCTCTGCTGAGCGGTATCCCTACGTGCGCGCGGTCAACATGCTCAAGAGAGAC 922  
 DB 11832 CCGAAGCGCTGCGCGCGCGCGGTGTCTCTGCGAGCGCGCGCATGCGCGAGAG 11773  
 QY 923 GCGTCTGCTGCGCGGAGCGACCATCATCTGTGTGCAAGCGCTTCCCGAATGAGCG 982

DB 11772 ACATCGCGCGGCTGCGGAGACTGCTGCTGCGGAGCGCGGCGGAGACTACCG 11713  
 QY 983 GCGCGACTGTGATGAAGTGTGCGCGCGGAGCGCG 1015  
 DB 11712 CCGCGCACCGCGGAGAGCGCGCGCGCGCGCGCTG 11680

Search completed: May 11, 2003, 01:57:18  
 Job time: 1433 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_12p model

Run on: May 11, 2003, 12:09:25 ; Search time 85 Seconds

(Without alignments) 8523.078 Million cell updates/sec

Title: US-10-009-782-1

Sequence: 1 gattccactgacgcgcgga.....ccctgacgacgagaagcct 1758

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus\_12p\_model -DEV=xlp  
-O=/gen2.1/USPRO/spool/US10009782/runat.07052003\_122517\_23135/app\_query.fasta.1.1927  
-DB=SPREMBL\_21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsun62 -TRANS=human40.ccl -LIST=45  
-LOCAL=IGM-200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MTELEN=0 -MAXLEN=2000000000  
-USER=US10009782.ecgn.1.1.99.gunat.07052003\_122517\_23135 -NCPU=6 -ICPU=3  
-NO\_XIPYX -NO\_XMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120  
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-GAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL\_21:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mmc:  
8: sp\_organella:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_virus:  
16: sp\_bacteriopl:  
17: sp\_archaeopl:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	2216	67.2	484 2	Q9AGH8 alcaigenes

Result No.	Score	Query Match Length	DB ID	Description
2	1064	32.3	494 16	Q8XGN5 ralsstonia s
3	942	28.6	526 17	Q9VZD3 pyrococcus
4	656.5	19.9	536 16	Q93JK8 streptomyces
5	407	12.3	581 12	Q9A2U6 calobacter
6	368	11.2	608 12	Q9Q5K9 herpesvirus
7	339	10.3	890 12	Q8UEZ1 ceratophthalic
8	338.5	10.3	529 12	P94309 epstein-bar
9	337	10.2	924 12	Q99307 pseudovirus
10	331	10.0	3084 12	Q8U211 ceratophthalic
11	322	9.9	608 12	Q9Q5K9 herpesvirus
12	318	9.8	890 12	Q8UEZ1 ceratophthalic
13	307	9.3	680 12	Q52747 ruminoecoc
14	306	9.3	586 6	Q95UD1 sus scrofa
15	299	9.2	706 12	Q41972 murid herpe
16	299	9.2	727 12	Q41973 murid herpe
17	299	9.2	774 12	Q41971 murid herpe
18	296	9.1	1791 4	Q60382 homo sapien
19	295	9.1	2752 4	Q9U035 homo sapien
20	295	9.1	1275 4	Q9U036 homo sapien
21	295	9.1	1783 4	Q15038 homo sapien
22	293	8.9	676 6	Q95JDO sus scrofa
23	292.5	8.9	511 6	Q95JDO sus scrofa
24	288.5	8.7	1325 5	Q9BKV7 leishmania
25	286.5	8.7	527 2	Q53913 streptomyc
26	286.5	8.7	1953 5	Q9B1T7 nephila mad
27	285.5	8.6	1404 4	Q92954 homo sapien
28	284.5	8.6	1315 10	Q9SPM0 zea mays (m
29	284	8.6	1404 4	Q9BX49 homo sapien
30	284	8.6	3247 12	Q65553 homo sapien
31	280	8.5	406 2	Q8VPMO micrococ
32	277.5	8.4	494 17	Q9HNR5 halobacteri
33	274.5	8.3	1453 11	Q63079 rattus norv
34	272.5	8.3	1414 5	Q26634 strongyloce
35	272	8.4	1262 4	Q9U040 homo sapien
36	272	8.4	2295 4	Q9U040 homo sapien
37	271	8.2	787 2	Q8VPMO micrococ
38	271	8.2	956 4	Q9U039 homo sapien
39	271	8.2	1262 4	Q9U040 homo sapien
40	271	8.2	2752 4	Q9U035 homo sapien
41	270.5	8.2	507 13	Q13028 boreogadus
42	270	8.2	1461 4	Q76045 homo sapien
43	268.5	8.3	601 5	Q62185 caenorhabd
44	268.5	8.1	2296 4	Q9U048 homo sapien
45	267.5	8.1	1445 13	Q93251 rana catesb

## ALIGNMENTS

RESULT 1  
ID Q9AGH8 PRELIMINARY; PRT: 484 AA.

AC 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE D-aminocyclase (EC 3.5.1.81).  
OS Alcaigenes faecalis.  
OC Bacteria; Proteobacteria; beta subdivision; Alcaigenaceae;  
OC Alcaigenes.  
OX NCBI-TaxID=511;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hsu C.S., Lai W.L., Chang W.W., Yang Y.B., Tsai Y.C.;  
RT "Gene cloning, nucleotide sequencing of Alcaigenes faecalis DAI D-  
RT aminocyclase";  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF332546; AAK15530.1; -  
KW Hydrolyase.  
SQ SEQUENCE 484 AA; 52000 MW; 634119A6CB75155E CRC64;

## Alignment Scores:

Pred. No.: 4.96e-104 Length: 484  
Score: 2216.00 Matches: 416

Percent Similarity: 93.80%  
 Best Local Similarity: 85.95%  
 Query Match: 67.17%

Conservative: 38  
 Mismatches: 30  
 Indels: 0  
 Gaps: 0

US-10-009-782-1 (1-1758) x Q9AGH8 (1-484)

```

QY 34 ATGTCCCAATCCGATTCCTCCAGCCCTTGCTGCGGGGCGGCGACCCATCGAC 93
Db 1 MetSerInProApLaThrProPheAspTyrIleLeuSerGlyGlyThrValIleAsp 20
QY 94 GCGGACCAACACCCGGGGGCGCGCGCGACCTGGCGGTGGCGGCGACCCATCGCGCC 153
Db 21 GlyThrAspAlaProGlyArgLeuAlaAspValGlyValArgGlyAspArgIleAla 40
QY 154 ATGGGCGATGTGGAGCGCGCGCGACACCCGGGTGCACCTGTGCGGCGCTGGTGC 213
Db 41 ValGlyAspLeuSerAlaSerAlaArgValArgIleAspValAlaGlyValVal 60
QY 214 GCGCGCGGCTTCATCGACTCGCACACCCAGACAGACAACTAGCTGTCAGGCGTGC 273
Db 61 SerProGlyPheIleAspSerHisThrHisAspAspAsnTyrIleuLeuLysHisAsp 80
QY 274 ATACGCCCAAGATCTCGCAGGGCGCTGACACGCGTGCAGGGGCAATGGCGCATAGC 333
Db 81 MetThrProLysIleSerGlnGlyValThrThrValValThrGlyLysnGlyIleSer 100
QY 334 CTGGCGCGGCTGGCGGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 333
Db 101 LeuAlaProLeuAlaHisAlaAsnProProAlaProLeuAspLeuAspGlnGly 120
QY 394 TCTTACCGTTTCGAGCGCTTCGCGGCTTCGCGGCTTCGCGGCGCGCGCGCGCG 453
Db 121 SerPheArgPheAlaArgPheSerAspTyrIleuGlnAlaLeuArgAlaAlaProProAla 140
QY 454 GTCAAGCGCGCGCTGTGTGGTGGCGCATTAAGCGTCCGCGCGCGTCAATGCGGACTTG 513
Db 141 ValAsnAlaAlaLysMetValGlyHisSerThrLeuArgAlaAlaValMetProAspLeu 160
QY 514 CACCGCGCGCGCACCGCACGAGAAATCGCGGCGATGCGGAGACTGGCGGAGGAGCGAG 573
Db 161 ArgArgGlnAlaThrAlaAspGluIleGlnAlaMetGlnAlaLeuAlaAspAlaLeu 180
QY 574 GCGAGGCGCGCATGCGGCTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
Db 181 AlaSerGlyAlaIleGlyIleSerThrGlyAlaPheTyrProProAlaAlaHisAlaSer 200
QY 634 ACCGAGAGATCATCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
Db 201 ThrGlnGluIleIleGlnValLysArgProLeuIleThrHisGlyGlyValTyrAlaThr 220
QY 694 CACATGCGCGGAGGAGGCGAGACATCGTGGCGCGCGCTGGAGGAAACCTTCGCGATCGGC 753
Db 221 HisMetArgAspGlnGlyGlnIstIleValGlnAlaLeuGlnGlnThrPheArgIleGly 240
QY 754 GCGGAGCTGGAGCTGCGCGGTGATCTGCACACAGATCATGGCGCGCGCGCGCGCG 813
Db 241 ArgGlnLeuAspValProValIleSerHisLysValMetGlyLysLeuAsnPhe 260
QY 814 GCGCGCTGGCGGAGACCTGCGCGTGTATGAGGCGCGCGCGCGCGCGCGCGCGCG 873
Db 261 GlyArgSerLysGlnThrLeuAlaLeuIleGlnAlaAlaMetAlaSerGlnAspValSer 280
QY 874 CTGACGCGTATCCCTAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
Db 281 LeuAspAlaTyrProTyrValAlaLysSerThrMetLeuLysGlnAspArgValAlaLeu 300
QY 934 GCGGAGCGCACCATCATCACTGTGTGCAAGCCCTTCCCGAAGTAGGCGCGCGCGACTG 993
Db 301 AlaGlyArgThrLeuIleThrTyrProLysArgProTyrProGlnLeuSerGlyArgAspLeu 320
QY 994 GATGAAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1053

```

```

Db 321 GluGlnIleAlaAlaGlnArgGlyLysSerLysTyrAspValValProGlnLeuGlnPro 340
QY 1054 GCGGCGCGCATCTACTCATGATGAGCAACCCGAGCGGCGCATCTCGGGCTTGGC 1113
Db 341 AlaGlyAlaIleTyrPheMetMetAspGlnProAspArgGlnAlaArgIleAlaAspGly 360
QY 1114 CCGAGCATGATCGGCTTCGAGCGCGCTCCGCGACGAGAGCGCGCGCGCATCCGCGCTTGG 1173
Db 361 ProThrMetIleGlySerAspGlyLeuProHisAspAlaProHisAspProHisAsp 380
QY 1174 GCGACCTCCCGGGGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1233
Db 381 GlyThrPheProLysValLeuGlnLysIstYSerArgAspLeuLysLeuPheProLeuGln 400
QY 1234 ACGGCGGTATGAGAATGACCGCGCTGACCGCGCGCGCGCGCTTCGCGGCGGCGGCGG 1293
Db 401 ThrAlaValTyrPheMetThrGlyLeuThrAlaAlaLysPheIleuAlaGlnValGly 420
QY 1294 CAGCTGAGCGCGGCTACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 1353
Db 421 GlnValGlnProGlyTyrTyrAlaAspLeuValValPheAspProAlaThrValAlaAsp 440
QY 1354 ACCGCGACCTTCGAGACCCCTTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413
Db 441 SerAlaThrPheGlnIstProThrGlnAlaGlnAlaGlyIleHisSerValTyrValAsn 460
QY 1414 GCGCGCGCGGCTTCGAGAGAGCGGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCA 1473
Db 461 GlyAlaAlaValTyrPheLysPheGlnIstYSerArgGlnHisAlaGlyArgValAlaAsn 480
QY 1474 CCGAGCGCGCGCG 1485
Db 481 ArgAlaGlyAla 484

```

## RESULT 2

OBXON5 PRELIMINARY: PRT: 494 AA.

```

ID 08XON5:
AC 08XON5:
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Probable N-acyl-D-glutamate deacylase protein (EC 3.5.1.82).
GN NDEP OR RSP187 OR RS03131.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID-305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM11000;
RX MEDLINE-21681879; PubMed-11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brotlier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Koisan A., Robert C., Saudin W., Schlex T.,
RA Stulier P., Thebaud P., Whalen M., Winkler P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AL646083; CAD18338.1;
KW Hydrolase; Plasmid; Complete proteome.
SQ SEQUENCE 494 AA; 53186 MW; 1C8A40B8B3EEC7C5 CRC64;

```

## Alignment Scores:

Pred. No.: 4.67e-46 Length: 494  
 Score: 1064.00 Matches: 229  
 Percent Similarity: 59.88% Conservat: 74  
 Best Local Similarity: 45.26% Mismatches: 183  
 Query Match: 32.25% Indels: 20  
 Gaps: 6

US-10-009-782-1 (1-1758) x 08XON5 (1-494)

OY	52	CAGCGCTTCAGCTCGCTCGCGGGGGGACCCCTATCGACGGCAGCAACCCCGGG	111
Db	2	GlnGlnTyrAspSerValIleAlaGlyValArgIleValAspIlySerIlyHsGluPro	21
OY	112	CGCGCGGCC-----GACGTGGGGGTGGCGGGGACCGACGATCGCGCCATCGGCATGTG	165
Db	22	GluSerThrLeuPheAspValAlaIleAlaAspGlyArgIleAlaIleAla---Thr	40
OY	166	TGGAGCGCGCGCGG-----CACACCCGGGTGACGTGTGGGGCCCTGTGGTGGCGG	216
Db	41	SerAspSerThrAlaTrpLeuGlyAspGluValAspIlyValAlaGlyAlaTrpValLeuAla	60
OY	217	CCCGGCTTCATCGATCGACACCCAGACAGACACTACCTCGCGGGGTGGCGACATG	276
Db	61	ProGlyPheIleAspValHsIleThrHsAspAspThrAsnValIleArgHyrProAspMet	80
OY	277	ACGGCCCAAGATCTGGCAGGGCGTCACACGCTGTGTACAGGGCAATTGGCGCATCACTGG	336
Db	81	LeuProIyValSerGlnGlyIleThrValIleValAlaGlyAsnGlyIleSerAla	100
OY	337	GGCGCGCGTGGCG---CACGGCAACCGCGCGCGCCCTGGACCTGTGTGGACGAGCGGC	393
Db	101	SerProAlaThrLeuAlaGlyGluProProAspProMetAsnLeuGlyProAlaGly	120
OY	394	TCTTACCGTTCGAGCGCTTCGCCGACATCTGTGACCGCGTGTGGCGGCACCGCGCGGCC	453
Db	121	AlaPheArgTyrProThrPheAlaSerTyrValGluAlaAlaValGluAlaIleAlaArgProAla	140
OY	454	GTCACAGCGCGCTGTANGGGGGGCGACATCAACGCTGGCGCGCGCGGTCAATGGCGGACTGG	513
Db	141	ValAsnValAlaAlaLeuValGlyHsThrAlaLeuArgTrpAsnHsIleuAspArgLeu	160
OY	514	CAGCGCGCGCGCCACGACGAGGAATCGCGCCATCGCGGACCTGGCGAGGAAGCGCATG	573
Db	161	AspArgProAlaThrAlaProGluValGluAlaIleAlaArgAlaGluLeuArgGluAlaLeu	180
OY	574	GCGACGGCGGCATCGGCATTTGTGACCGGGCGCTTACCGCGCGCGCGCGCGCGCAC	633
Db	181	AspHsIleGlyAlaLeuGlyLeuSerThrGlyLeuAlaIleTyrAlaAsnAlaHsAlaAlaAla	200
OY	634	ACCGAAGAGATCATCGAGGTGGCGCGCCCTAGAGCGCGCATCGGCATTCACCGCAC	693
Db	201	ThrGluGluValMetSerLeuAlaGluProLeuAlaGluAlaIleValLeuTyrThr	220
OY	694	CACATCGCGCACGAAGCGCGACACATCGTGTGGCGCGCTGGAGGAACCTTCCGCATCGCG	753
Db	221	HsLeuArgThrGluPheAlaAlaIleLeuGluAlaMetAspIleValIleTyrArgValGly	240
OY	754	CGCGAGTGGAGCTGGCGGGTGGATGTGTCACACAGAGTCACTGGGCGGACGCGCATTTGC	813
Db	241	ArgHsAlaArgValProIleValIleSerHsIleuIyScyAlaIleValAlaAsnTrp	260
OY	814	GGCGCTCGCGCGAGAGCTCGCGCTGATGAGAGCGCGCATGGCGCGCGAGACGCTCTCG	873
Db	261	GlyTrpAlaGlyGluValIleuGluSerIleGluAlaAlaGluArgTyrGlnProValGly	280
OY	874	CTGAGACGGTATCCCTACGTGGCGCGGCTCCACATGCTCAAGAGACCGCGTGTCTGTG	933
Db	281	CysAspCysTyrProTyrThrAlaSerSerThrIleuAspIleuIyGlnValAlaThrAsp	300
OY	934	GCGCGAGACACATCATCACCTGGTGGACACCTTCGCCGACGTGAGCGCGGACGCTG	993
Db	301	GluTrpAspIleValIleThrTrpSerAspTrpHsIleProGluMetAlaGlyIleThrIle	320
OY	994	GATCAAGTGGCGCGCGAGCGCGGCAAAATCCAAATGACATGACATGGTGGCGCGAGCTGCACGG	1053
Db	321	LysAlaIleAlaAlaAlaAlaTrpGlyThrAspIleuAspIleAlaIleAlaArgTrpLeuGlnPro	340
OY	1054	GCGCGCGCATCTTCATCATGTGAGAGAACCGGACGTGACGACATCCGCGGCTTCGGCG	1113
Db	341	AlaGlyAlaValIlyThrHsAsnMetAlaAlaGluInsPheuSprArgIleIleAlaHsPro	360

QY	1114	CGGACCATGATGGCTCCGACGGCTCGCGGACAGAGAGCGGCCGATCCGGCCCTGTGG	1173
DG	361	AlaThrValIleGlySerAspGlyLeuProAsnAspProGlnProHisProArgLeuTrp	380
QY	1174	GGCACCCTCCCGGGGCTGGGGGACATAGCGCGGACCTGGGCGCTCTTCCCGTGGAG	1233
DG	381	GlyAlaPheProHisValLeuAlaHisTyrSerArgGluArgGlnIleuLeuSerLeuAla	400
QY	1234	ACGGCGGATATGGAAAGAACCGGCCCTGACCCCGCGCCGCTTGGGCGCTGGCCGGCGGG	1293
DG	401	AlaAlaValArgLysMetTrpGlyGlnSerAlaGluArgPheGlyLeuAlaGluArgGly	420
QY	1294	CAGCTGAGGACCGGGCTACTTCGCGGACCTGGTGGTGTGGACCCCGCCAGCTGGCCAT	1353
DG	421	LeuValArgGluGlyTrpTrpAlaAspLeuValLeuPheAspProAspPheValArgAsp	440
QY	1354	ACGGCCACCTTCGAAACACCTACCGGACGGCGCCGCGGCATCCATTCGTGATCTAAC	1413
DG	441	LeuAlaThrPheThrAspProLysGlnProAlaIleAlaGlyIleGluValValTrpValAsn	460
QY	1414	GGCGCGCGGCTGGCGAAAGAGCGGGGTCACCGGACACAGACGGCGCGCTGGTCGA	1473
DG	461	GlyAlaLeuSerTyrArgAspGlyAlaValGlnProArgArgAlaGlyArgPheLeu---	479
QY	1474	CGCAGCGCGCTGAGCCCGCGGCCAGCCCTTACAAATCCGGCGTGAACGGCGGGCGTGC	1533
DG	480	-----ArgArgGlyPro-----ArgAlaHisAlaA	487
QY	1534	CGGCCCTCCCAACCGCG	1551
DG	488	GlnProAlaProGluIleu	493

RESULT 3			
ID	Q9VZD3	PRELIMINARY;	PRT; 526 AA.
AC	Q9VZD3;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	D-aminoacylase (Aspartate, glutamate etc).		
GN	NDAD OR PAB0090.		
OS	Pyrococcus abyssi.		
OC	Archaea; Euryarchaeota; Thermococci; Thermococcaceae; Thermococcaceae;		
OC	Pyrococcus.		
OX	NCBI_TaxID-29292;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-ORSAY;		
RA	Hellig R.;		
RT	"Pyrococcus abyssi genome sequence: Insights into archaeal chromosome		
RT	structure and evolution."		
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ248283; CAB4965.1;		
DR	InterPro; IPR002833; Adenine_dcam.		
DR	Pfam; PF01979; Adenine_dcam; 1.		
KW	Complete proteome.		
QC	SEQUENCE 526 AA; 58790 MW; 705B1B7B281915A0 CRC64;		

Alignment Scores:	
Pred. No.:	6,49e-40
Score:	94.00
Percent Similarity:	55.56%
Best Local Similarity:	40.11%
Query Match:	28.55%
DB:	17
Length:	526
Matches:	213
Conservative:	82
Mismatches:	172
Indels:	64
Gaps:	17

US-10-009-782-1 (1-1758) x Q9V2D3 (1-526)

QY 58 TTTCACCTGCGTGGCGGGGGCGGACCCGATCGAGCGAGCAACACCCCGGGCGCGGC 117

Db 4 TTTTACCTGCGTGGCGGGGGCGGACCCGATCGAGCGAGCAACACCCCGGGCGCGGC 117

QY 118 GCCACCTGGGCGTGGCGGGCGGACCCGATCGAGCGAGCAACACCCCGGGCGCGGC 177







0Y	58	TTGACGCTGCTGCTGGGGGGGACGACCTATCTGACGGGACCAACCCCGGGGGCGC	117
0Y	59	:::	118
Db	6	TYRASPLeuValLeuLdaTgSerGlyValValAPheASPLeuAlaGlyLeuProGlyPheGln	25
0Y	118	GCGGACCTGGGGGTGCGGGGGGACCGGATGCGCGGCGCATTCGGGAGCGCGCC	177
0Y	119	:::	178
Db	26	AlaASPValGlyValAlaTgSPeGlyValIleValAlaValGlyLys---GlyLeuAlaAla	44
0Y	178	GCGGACACCGGGGTGACGTGTGCGGCGCTGTGTTCGCGCGCGGCTTCATGACGTGAC	237
0Y	179	:::	238
Db	45	GlyAlaGlnGlnIleASPAlaLargGlyValArgIleValIhrProGlyPheValASPLeuHis	64
0Y	238	ACCGACGACGACAACTACTCTCTAGGCGTGGCGGACATGACGCCCAAGATCTGGAGGC	297
0Y	239	:::	298
Db	65	ThrIhrTYRASPLeuGlnAlaIhrTrpSerGlyLeuGlnGlyProSerSerGlyHisGly	84
0Y	298	GTCACACGAGGGGTGACGGGCAATTGGGGCATTCAGCCTGGCGCGC	342
0Y	299	:::	343
Db	85	ValIhrTrpValValMetGlyAsnCysGlyValGlyPheAlaProCysArgProASP	104
0Y	343	-----CTGGCGGACGCGCAACCGCGCGCGCGCGCGCGCGCTGGACGTGGAC	384
0Y	344	:::	385
Db	105	HisASPArgLeuIleIhrGleuMetGlnGlyValGlnAspIleProPheProValIhr	124
0Y	385	GAAAGCGGCTCTTACCGCTTTCGAGCGCTTGGCGGACATCGGACGCGTTCGGGGCACG	444
0Y	386	:::	445
Db	125	GlnGlyLeuProTrpAlaTrpGlyLeuSerPheASPArgLeuAspPheLeuAlaGlyArg	144
0Y	445	CCGGGCGCGCCCAACCGCGCTGTATGTTGGTGGCCATTCACGCTGGCGCGCGCGGTGATG	504
0Y	446	:::	505
Db	145	AlaPheASPValASPValGlyValGlnLeuProHisAlaIhrLeuArgValTYRValMet	164
0Y	505	CCGGACCTTG-----CAGCGCGCGCGGACGACGACGAGAAATCGCGGACATCGGGAC	555
0Y	506	:::	556
Db	165	GlyASPArgGlyValASPArgGlyLeuProAlaIhrASPAlaSPLeuAlaIhrAlaAla	184
0Y	556	CTGGCGGACGAAAGCCATGGCGGCGGCGGCGCATTCGGCATTCGACCGGCGCTTCTAC--	612
0Y	557	:::	613
Db	185	IleAlaArgArgAlaValAlaGlnIleGlnIleValIhrPheSerThrSerArgIhrLeuAsn	204
0Y	613	-----CGCGCGCGCGCGCGCGCGCGGACGACGACGAGAAAGATCANTC	648
0Y	614	:::	649
Db	205	HisArgTrpSerASPArgGlyLeuProIhrProIhrLeuIhrAlaGlyLysArgIhr	224
0Y	649	GAGGTGTGCGCGCGCTGAGCGCGCATGGCGGCGCATTCACGCCACACCGACGCGGAC---	705
0Y	650	:::	706
Db	225	GlyIleAlaLeuGlyLeuAlaIhrAlaIhrGlyGlyValLeuGlnValIhrSerASPhe	244
0Y	706	---GAGGCGGACCATCTGTGGCGCGCGCTGGAGGAAACCTC---CGCATCGCGCGCGAG	759
0Y	707	:::	760
Db	245	ValASPArgAlaAlaGlnIleAlaIhrMetLeuArgArgIleValGlnIhrSerGlyArgPro	264
0Y	760	CTGGACGTGGCGGTGGGTGATCTG-----	783
0Y	761	:::	784
Db	265	LeuSerPheSerLeuValGlnSerProGlyAlaProGlnGlyTrpArgGlyLeuAsp	284
0Y	784	-----CACCAAGCATGGCGCGCGCC	807
0Y	785	:::	808
Db	285	GlyValAlaAlaAlaValASPAlaGlyLeuProMetLeuAlaGlnValCysGlyArgPro	304
0Y	808	-----AATTTCGGC-----	816
0Y	809	:::	817
Db	305	ValGlyValLeuPheGlyLeuGlnIleuThrLeuAsnProPheSerGlnAsnProValAlaPhe	324
0Y	817	CGCTCGCGCGGACGCTGCGCGGTGATGAGGCGCGCGCATGGCGCGGACGCTGTGCGTG	876
0Y	818	:::	877
Db	325	AlaGlnLeuLysASPArgProLeuAlaASPLeuValAlaAlaLeuSerASPProAlaPhe	344
0Y	877	GACGCGTATCCCTACGTGGCGCGCGGCGCACCATGCTCAAGCAGAGC-----GCGCGG	927
0Y	878	:::	928
Db	345	ArgAla-----ArgLeuLeuAlaHisASPLeuASPAlaAlaArgGly	357

QY	928	CTGCTGGCCGGACGACCATCATCTACTGTGGCAAGCCCTTCCCGCACTAGACGGGCGC	987
Db	356	ProPhelaAGlySerAlaLeuAlaValatPaspasIndeuIrrPro---MetGlyVal	376
QY	988	GACCTGAT-----	1011
Db	377	AspProAspTyrGluProThrAlaAspLysThrValAlaAlaLeuAlaAlaArgGluGly	396
QY	1018	AAATCAAGTACGACGTGGTCCCGAGCTGCAGCCGCC-----GGCGCCATC	1065
Db	397	ArgAspProAlaAlaValAlaLeuAspAlaMetLeuAlaIatrgspgLyIatrgLyMetLeu	416
QY	1066	TACTTC---AGATGGACGAACCCGAGCTGCAG-----CGCATCTCGGCG	1107
Db	417	TyrHisProPheLeuAsnTyrAlaAspGlySerLeuAspProSerPheAlaMetLeuSer	436
QY	1108	TTCCGCCCGCAGCATGATCGGC---TCCGACGGCGTGGCGAC-----GAC	1144
Db	437	HisArgAspThrValProGlyLeuSerAspGlyGlyAlaHisValGlyMetLeuLysAsp	456
QY	1150	GAGCGCCCGCATCCGCCCTGTGGGCGACCTTCCCGCGGGT---CTGGGCGCATATGCG	1206
Db	457	-----GlySerPheProThrSerAsnLeuLeuHisThrPthr	468
QY	1207	CGCGACCTG-----GGCTGTTCCTCCCTGAGAACGGCGGTATGAAATGACCGGC	1257
Db	469	ArgAspArgTyrThrArgGlyProArgLeuPheLeuGlyIlePheLeuGlyIleMetLeuAlaArgIndSerArg	488
QY	1258	CTGACCGCGCGCGGCGCTTGGCGCTGGCGGGCGGGGACGTGCAGCGCGGATGACTTCGCG	1311
Db	469	AspThrAlaGlnAlaValGlyLeuLeuAspArgGlyLeuLeuAlaProGlyIlyIatrgAla	508
QY	1318	GACCTGTGTGTTCGACCCGCGCACGGTG-----GCCGATACCGCACCTTCGAA	1366
Db	509	AspLeuAsnValIleAspTyrAlaGlyLeuArgLeuGlnAlaProGlnValAlaIatrgAsp	528
QY	1369	CACCT-----ACCGAGCGGGCGCGCGGACATCCATTCCTCGTACGAC	1411
Db	529	LeuProAlaGlyIatrgAlaGlyLeuThrGlnIatrgAlaHisGlyTyrValAlaThrIleVal	548
QY	1411	AACGGCGCGCGGCTGTGGCAAGACAGAGCGCTTACCGCGCAGACATCCCGCGCGTGTCT	1470
Db	549	AlaGlyValThrGlnIatrgAspGlyGluProThrGlyAlaLeuProGlyIatrgLeu---	567
QY	1471	GCAGGACAGCGCGGCTGAGCCGCGCGCACGCTTACAAATCCGGCGGTGAACGGGCGCGC	1533
Db	568	-----ValArgGlyAlaGln	572
QY	1531	TGCCGCCCGCTC	1542
Db	573	SerAlaProLeu	576
RESULT 6			
Q905K9		PRELIMINARY;	PRT: 608 AA.
AC	0905K9		
DT	01-MAY-2000	(TREMBlrel. 13, Created)	
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)	
DT	01-MAY-2000	(TREMBlrel. 13, Last annotation update)	
DE	NTR.		
OS	Herpesvirus paplo.		
OC	Vituses; dsDNA vituses, no RNA stage; Herpesviridae;		
CC	Gammaherpesvirinae; Lymphocryptovirus.		
OX	NCBI_TaxID=10394;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BABOON LYMPHOCRYPTOVIRUS BA65;		
RA	Zong J.-C., Ryan J., Ling P.D., Loeb D.D., Pagano J.S., Hayward S.D.,		
RL	Hayward G.S.		
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.		
QR	EMBL; AF000364; AAF23950.1;		
QK	SEQUENCE 608 AA; 60458 MW; 48782ACBA0A029627 CRC64;		



DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE LF3.  
 OS cercopithicine herpesvirus 15.  
 CC viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 OX NCBI\_TaxID=104228;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LCL8664;  
 RC MEDLINE=97048062; PubMed=8892903;  
 RA Franken M., Devergne O., Rosenzweig M., Annis B., Kieff E., Wang F.;  
 RT "Comparative analysis identifies conserved tumor necrosis factor  
 RT receptor-associated factor 3 binding sites in the human and simian  
 RT Epstein-Barr virus oncogene LMP1.";  
 RL J. Virol. 70:7819-7826(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LCL8664;  
 RC MEDLINE=99412410; PubMed=10482645;  
 RA Rivallier P., Quink C., Wang F.;  
 RT "Strong selective pressure for evolution of an Epstein-Barr virus  
 RT LMP2B homologue in the rhesus lymphocryptovirus.";  
 RL J. Virol. 73:8867-8872(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LCL8664;  
 RC MEDLINE=20304984; PubMed=10846073;  
 RA Jiang H., Cho Y.G., Wang F.;  
 RT "Structural, functional, and genetic comparisons of Epstein-Barr virus  
 RT nuclear antigen 3A, 3B, and 3C homologues encoded by the rhesus  
 RT lymphocryptovirus.";  
 RL J. Virol. 74:5921-5932(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LCL8664;  
 RC MEDLINE=20440633; PubMed=10970361;  
 RA Rao P., Jiang H., Wang F.;  
 RT "Cloning of the rhesus lymphocryptovirus viral capsid antigen and  
 RT Epstein-Barr virus-encoded small RNA homologues and use in diagnosis  
 RT of acute and persistent infections.";  
 RL J. Clin. Microbiol. 38:3219-3225(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LCL8664;  
 RC MEDLINE=21602573; PubMed=11739708;  
 RA Rivallier P., Jiang H., Cho Y.-G., Quink C., Wang F.;  
 RT "Complete Nucleotide Sequence of the Rhesus Lymphocryptovirus: Genetic  
 RT Validation for an Epstein-Barr Virus Animal Model.";  
 RL J. Virol. 76:421-426(2002).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LCL8664;  
 RC Mognaddam A., Koch J., Annis B., Wang F.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LCL8664;  
 RC Mognaddam A., Annis B., Wang F.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LCL8664;  
 RC Rivallier P., Quink C., Wang F.;  
 RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LCL8664;  
 RC Jiang H., Wang F.;  
 RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [10]  
 RP SEQUENCE FROM N.A.

RC STRAIN-LCL8664;  
RA Rao P.V., Jang H., Wang F.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [11]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RA Rivaller P., Jang H., Cho Y.-G., Quirk C., Wang F.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY037858; AAK95470.1; --  
SQ SEQUENCE 890 AA; 95037 MW; 149077CB9D3B5574 CRC64;

Alignment Scores:

Pred. No.:	1,53e+09	Length:	890
Score:	339.00	Matches:	19
Percent Similarity:	35.69%	Conservative:	43
Best Local Similarity:	29.35%	Mismatches:	259
Query Match:	10.28%	Indels:	178
DB:	12	Gaps:	40

OS-10-009-7.82-1 (1-1758) x Q8UZE1 (1-890)

51 ccagcccttcgacctgcttcc--ggcgggacccctcatcgacggcgacacacccc 107

[illegible][illegible][illegible]

210 GGTCCGC-----CCGGCTCATCGACTCGCACACCCACGACGACAACTACCTGCTCAGGC 265

Db 88 GlyThrArgProProAlaProSerSerArgThrGlyGlnArgThrArgGlnArgProGly 107

266 GTCGCACATGACGCCAAGATCTGCAGGGCGTCAACCACGGTGTCAAGGGCAATTGCG 325

Db 108 HisPro-----ProProGluArgGlySerGlyProArgGlyThrArgProProAla 124

326 GCATCAGCCTGGCGCCGCTGGCGCAG-----CCAAACCGCGCCGCCCTGGACC 376

00 123 FROBESIA GILGIANA GRACIOLA PROBYNSIS PROPRO----- 142

[illegible][illegible][illegible]

0v      491      63A

Db . 178 GLYPROARQGLYThrARQPROProAlaProSerSerARAThrgIvgInArcthrARAGIn 197

521 -----CCGCCACCGAGGAAATCGCGGCATGCGGACCTGCGCGAGGAAG 568

Db 198 ArgProGlyHisProProGlyArgGlySerGlyProArgGlyThrArgPro----- 215

569 CCATGGCCAGCGGCCCATCGGCATTTCGACCGGGCGCCTTCTACCGC----- 616

216 -----ProAlaProSerSerArgThrglyGlnArg 225

CCGCCCCC  
GGGCCA 631

[illegible]

9  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525





```

QY 362 CCGCCCCCTGAGACCTGCTGAGCAAGCCGCTTACCGTTTCAGACGCTTCCGCACT 421
DB 111 ||||| ||||| |||||
QY 422 ACGTGGACCGCTTGGCGGCGCCAGCGCGCGCTCAACCGCTGTATGTGGCCATT 481
DB 123 -----ArgLeuGlyAlaHis----- 127
QY 482 CAACCTGCGCGCGCGGCGCATGCGGACCTTGCAGCGCGCGCCACCGAGCAAGATCG 541
DB 128 -----ProAlaArgAlaArgAlaArgHisLeu----- 122
QY 542 CCGGCATCGCGGACCTGCGCGGAGCAAGCCATGCGCGCGCGCGCATTCGCAATTCGACCG 601
DB 141 GAlaArgLeuGlyAlaArgGlyAlaArgGly-----ProArg 151
QY 602 GCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 661
DB 151 GAlaArgLeuAlaArgGlyAlaArgGlyAlaArgGlyAlaArgGlyAlaArgGlyAla 170
QY 662 CCGTGAAGCGCGCATGCGCGCGCATTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 721
DB 170 YLeu-----HisArgAlaArgAlaArgAlaArgAla 180
QY 722 TGGCGCGCGCTGAGCAACCTT-----CGGCATCG 751
DB 180 GAlaArgLeuAlaArgProGlyAlaArgHisGlyAlaArgProGlyAlaArgProGlyAla 200
QY 752 GCGCGGAGCTGAGCGCTGCGCGCTGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 811
DB 200 GProGlyAlaArgProGlyAlaArgHisProLeuArgGlyGlyGlyGlyGlyGlyGlyAla 220
QY 812 TCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 871
DB 220 AlaArgProLeuAlaArgAla-----ArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 235
QY 872 CCGTGAAGCGCGCTTCCCTACGCTGCG-----CGGCTTCCAGCATTCAGCAAGCGCGCG 925
DB 235 AlaArgGlyAlaArgHisProAlaArgLeuGlyAlaArgHisArgAlaArgProGlyAlaArg 255
QY 926 TCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 985
DB 255 GAlaAlaHisArgGlyAlaArgHisLeuArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 268
QY 986 GCGACCTGATGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1045
DB 269 -----ArgArgGlyAlaArgAla-----AspGlyCysArgAla 278
QY 1046 TCGACGCGCGCGCGCGCGCATCTACTTCATGATGAGCAAGCCGCGCGCGCGCGCGCG 1105
DB 278 AlaArgLeuAlaArgAlaArg-----ArgAlaArgAlaArgAlaArgGlyGlyGlyGly 292
QY 1106 CGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1141
DB 292 YLeuArgAlaArgGlyAlaArgGlyAlaArgHisArgProArgProArgGlyAlaArgGly 312
QY 1142 CCGACGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1201
DB 312 YAlaAlaArgLeuAlaArgAlaArgGlyAlaArgGlyAlaArgProAlaArgProAlaArg 332
QY 1202 A---TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1232
DB 332 CAlaArgHisArgProAlaArgProAlaArgProGlyAlaArgGlyAlaArgGlyAla 352
QY 1233 -----GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1288
DB 352 HisArgProGlyAlaArgLeuGlyAlaArgGlyAlaArgGlyAlaArgGlyAlaArgGlyAla 372
QY 1289 GCGCGGAGCTGCGCGCGCG-----GACTTCCGCGCGCGCGCGCGCGCGCGCGCG 1322
DB 372 AlaArgLeuHisGlyAlaArgGlyAlaArgHisArgProArgAlaArgAlaArgGlyAla 392

```

```

QY 1323 -GGTGGTGTTCGACCGCGCG-----CACGGTGGCGCGGATACCGCGCGCGCGCGCG 1372
DB 392 GAlaArgGlyProAlaArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 410
QY 1373 CTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1432
DB 410 AlaArgAlaArgAlaArgGlyAlaArgGlyAlaArgGlyAlaArgGlyAlaArgGlyAla 430
QY 1433 AGCAGCGCGTTCACG-----GGCGACGATTCGCGCGCGCGCGCGCGCGCGCGCGCG 1483
DB 430 HisArgAlaArgAlaArgAlaArgAlaArgAlaArgAlaArgAlaArgAlaArgAla 450
QY 1484 CTTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1543
DB 450 YLeuLeuAlaArgAlaArgAlaArgAlaArgAlaArgAlaArgAlaArgAlaArgAla 459
QY 1544 CAACCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1600
DB 460 -HisLeuAlaArgAlaArgAlaArgAlaArgAlaArgAlaArgAlaArgAlaArgAla 476
QY 1601 CCGATATGTTGGGCAAGAGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1656
DB 477 -----TrpThrValMetAlaThrArgLeuArgGlyAlaArgGlyAlaArgGlyAla 492
QY 1657 -----CGGAAATGACCTTCGCAAGACCTGTGCGCGCGCGCGCGCGCGCGCGCG 1689
DB 492 AlaAlaArgAlaArgProThrThrSerSerSerSerSerSerSerSerSerSerSer 505

RESULT 9
ID 099307 PRELIMINARY; PRT; 924 AA.
AC 099307;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-FEB-1997 (TREMBLrel. 02, Last annotation update)
DE L33 protein.
OS Epstein-Barr virus (strain Raj1) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10378;
RN [1]
RP SEQUENCE FROM N.A.
RA Parker B.D., Bankier A., Satchwell S., Bartell B., Farrell P.J.;
RL Virology 0:0-0(1990).
RT EMBL; M3547; AAA45896.1;
SQ SEQUENCE 924 AA; 94304 MW; 3C7CC68297F0E0FA CRC64;

Alignment Scores:
Pred. No.: 1.94e-09 Length: 924
Score: 337.00 Matches: 210
Percent Similarity: 34.58% Conservative: 30
Best Local Similarity: 30.28% Mismatches: 287
Query Match: 10.22% Indels: 168
Gaps: 31

US-10-009-782-1 (1-1758) x 099307 (1-924)
QY 45 CGATCCCGACCGCTTGCAGCT---CTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 101
DB 6 ArgGlyProCysLeuAlaArgProGlySerGlyLeuGlyAlaHisArgHisArgProArgGlySer 25
QY 102 -----CACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 131
DB 26 GAlaAlaArgProAlaArgProAlaArgProAlaArgProAlaArgProAlaArgProAla 44
QY 132 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 179
DB 45 ---ProGlyProArgGlyAlaArgGlyAlaArgGlyAlaArgGlyAlaArgGlyAlaArg 179
QY 180 -----GCACACCGCGCGT-----CGACGCTGCGCGCGCGCGCGCGCGCGCGCG 218
DB 64 ProAlaArgProAlaArgGlyAlaArgHisArgProAlaArgProGlyProGlyProGlyThr 83

```

```

OY 219 CGGCTTCATGACCTCGACACCCAGACCAACTACCTGCTCAGGGGTCGACATGAC 278
    |||||  ::  |||  |||||  |||  ::|||
Db 84 ArgLeu---GlnProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAsp 102
OY 279 GCCCAAGATGCTCCAGGGCTCACACACGTCGTCACGGGCAATTGCGCATACCTGCG 338
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 103 GlnHisProAlaAlaProAlaAlaProGlyProGlyProAlaThrArgLeuGlnProAla 122
OY 339 GCC-----GCTGGGCGACGCGCAACCCCGCCGCGCC----- 368
    ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 123 ThrProAlaGlySerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaPro 142
OY 369 -----CTGAGACCTGCTGAGACGAGCGGCTC-----TACCGTTTCAGCGCTT 413
    |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db 143 ArgAlaProGlyProGlyProAlaGlyHisArgLeuGlnProAlaThrProAlaGlySerGly 162
OY 414 CGCGGCACTACCTGGACGCGCTGGCGGCGACCGCGCGCGCTGACACCGCCCTGTATGCT 473
    |||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 163 AlaAlaAspProAlaAspProValGlyHis-ProAlaAlaProAlaGlyProGlyProG 182
OY 474 GGGGCTTCAGAGCTCGCGCGCGCGCGCTCATCGCGACCTGACCGCGCGCGCGCGACGA 533
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 182 LnpProAlaGlyThrArgLeuGlnProAlaThrProAlaGlySerGlyAlaAlaAspPro 200
OY 534 GGAATTCGCGGCGGCTCGGACCTGGCGCGGACGCGGCAAGCGCATGCGCGCGCATCGG 593
    ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 201 -----AlaAspProValGly-----HisProAlaAlaProAlaGlyAla 213
OY 594 TTCGACCGCGGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 213 ProGlyProAlaGlyThrArgLeuGlnProAlaThrPro-AlaGlySerGlyAlaAla 222
OY 642 -----GATCATCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 692
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 223 AspProAlaAspProValGlyHisProAlaAlaProAlaGlyAla-----ProGlyProG 250
OY 693 CCACATGCGCGCGACGAGCGGACACATCGTGGCGCGCGCGCGCGCGCGCGCGCGCATCG 752
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 251 ProAlaGlyThrArgLeuGlnProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAsp 270
OY 753 CGGAGAGCTGAGCGCGCGCGGTGATCTCGACCAACAGTCATGCGCGCA----- 803
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 271 Pro-----ValGlyHisProAlaAlaProAlaAlaProGlyProGlyProAlaG 286
OY 804 GCCCAATTCGCGCGCGCGCGCG-----CGAGACGCTGCGCGTGAATGAGAGCG 848
    ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 287 ThrArgLeuGlnProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAspProVal 306
OY 849 CGCCATGCGCGCGCGCGCGCGCTCGCTGCA-----CGCGTATCCCTACGTCGCGG 899
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 307 GlnHisProAlaAlaProAlaGlyAlaProGlyProGlyProAlaGlySerGlyAlaAla 326
OY 900 CTCACCACTGCTCAACGACGAGCGCG-----TGCCTGCTGCGCGGACGCGAC 944
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 326 LnpProAlaGlySerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaPro 346
OY 945 CATATCATCGCTGCTGCGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1004
    ||  ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 346 LnpProAlaGlyProGlyProAlaGlyHisArgLeuGlnProAlaThr-----ProAlaG 364
OY 1005 GCGCGAGCGCGCGCAATCAATACGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1058
    ||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 364 GlnSerGlyAlaAlaAspProAlaAspPro-ValGlyHisProAlaAlaProAlaGlyAlaPro 383
OY 1059 -----CGCATCTCACTTCATGATGAGACGACCGCGCGCGCGCGCGCGCGCGCG 1106
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 384 GlyProGlyProAlaGlyThrArgLeuGlnProAlaThrProAlaGlySer-----Gly 400
OY 1107 GTTCGCGCGCGCGCATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1166
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 401 AlaAlaAspProAlaAspProValGlyHisProAlaAlaProAlaGlyAlaProGlyProG 420
OY 1167 CCGTGGGCGCGACCTT----- 1181

```

```

Db 421 ProAlaGlyThrArgLeuGlnProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAsp 440
OY 1182 CCGCGCGGCTGCTGGGCGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1241
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 441 ProValGlyHisProAlaAlaProAlaGlyAlaProGlyProGlyPro----- 455
OY 1242 ATGGAAGATGACCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1301
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 456 -----ArgThrArgLeuGlnProAlaThrProAlaThrProAlaGlySerGly 468
OY 1302 GCGCGGCTACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1360
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 469 AlaAlaAspProAlaAspProValGlyHisProAlaAlaProAlaGlyAlaProGlyProG 488
OY 1361 CCTCGAAGACCCCA----- 1375
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 489 ProAlaGlyThrArgLeuGlnProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAsp 508
OY 1376 -----CCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1423
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 509 ProValGlyHisProAlaAlaProAlaGlyAlaProGlyProGlyProAlaGlyHisArg 528
OY 1424 TCTGGCAAGACGAGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1465
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 529 ProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAspProValGlyHisProAla 548
OY 1466 TGCACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1519
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 549 AlaProAlaGlyAlaProGlyProGlyProAlaGlyThrArgLeuGlnProAlaThrPro-AlaG 568
OY 1520 ACGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1564
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 568 GlnSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProAlaGlyAlaPro 587
OY 1565 ACATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1624
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 588 -----GlyProGlyProAlaGlyThrArgLeuGlnProAlaThrProAlaGlySerGlyAlaAla 606
OY 1625 T-----GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1664
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 606 AspProAlaAspProValGlyHisProAlaAlaProAlaGlyAlaProGlyProGlyProAla 626
OY 1665 -----GACCTGCAAGACCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1714
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 626 GlnThrArgLeuGlnProAlaThrProAlaGlySerGlyAlaAlaAspProAlaAspProVal 646
OY 1715 CCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1742
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 646 GlnHisProAlaAlaProAlaGlyAlaPro 655

```

RESULT 10

08U211 PRELIMINARY; PRT: 3084 AA.

AC 08U211:

DT 01-VAR-2002 (TREMBLrel. 20, Created)

DT 01-VAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE UL36 protein

OS Pseudorabies virus (strain Kaplan) (PRV).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Varicellovirinae.

OX NCBI\_TaxID=33703;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KAPLAN.

RA Klupp B.G., Fuchs W., Granzow H., Nixdorf R., Wettenleiter T.C.,

RT "The Pseudorabies virus UL36 tegument protein physically interacts

RT with the UL37 protein."

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ422133; CADI9511.1; -

DR InterPro; IPR005210; Herpes\_UL36.

DR Pfam; PF03586; Herpes\_UL36; 1.

**SEQ**    **SEQUENCE**    :    **3084 AA**;    **324401 MW**;    **ECAD9E1E3DC22D1A CRC64**;

**Alignment Scores:**

Pred. No.:	4,53e+09	Length:	308
Score:	331.00	Matches:	187
Percent Similarity:	32.72%	Conservative:	45
Best Local Similarity:	26.38%	Mismatches:	269
Query Match:	10.03%	Indels:	209
DB:	12	Gaps:	29

US-10-009-782-1 (1-1758) x Q8U211 (1-3084)

QY	14	TCGCGGAGAGAGATTCCATGTCCCAATCCGAT	-----CCGAGCCCTCGAC	64
Db	2253	SeirleuYalAalAgtlieGluProProProProValIlaAsnProProLyProAlaThr		2272
QY	65	TGCTGCTCGCGGGCGGCACCTCATCGAGCGGCAACACCCGGGGCGCGCCGCGAC		124
Db	2273	-----AlaProGlnAlaProGluThrAlaProGlu		2288
QY	125	TGGGCGTGGCGGGGACCCGATCGCGCCATCGGCACTGTGCGAC		172
Db	2283	-----AlaProGlnAlaGlnProProAlaAlaAlaAlaProThrProGlnProGln		2299
QY	173	-----CCGCGCGCACACCGGGGTGAGCTGCGGGCTGGGGTGGCGGGCGGT		223
Db	2300	GLPProProProProProGlnProProProSerAlaGlnAlaProProAlaGlnProPro		2319
QY	224	TCATCGACTGCAACCCACGACAGACACTACCTGTCTGAGCGTGCAGCATGAGCCCA		283
Db	2320	AlaGlnProAlaThrAlaAlaAlaThrThrAlaProLyAlaThrProGlnThrGlnPro		2339
QY	284	AGATCTGCAAGGCGCTCACCCAGCGTGTGACGGGCAATTGCGGATCAGCTGCGCGCG		343
Db	2340	ProThrAlaAla	-----	2343
QY	344	TGGGCGGCGCAACCGCGCGCCCGCCCTGACCTGTGACGACAGCGGCTTTAC	-----	400
Db	2344	-----GlnThrGlnThrAlaProProProProSerAlaAlaThrAlaAlaAlaGlnAlaPro		2362
QY	401	-----GTTTCGAGCCTTCGCGGACTTACCTGAGCGCGTTCGGGCGACCGCGGGCG		454
Db	2363	ProGlnProProSerSerGlnProAlaAlaAlaProAlaGlyAlaProAlaPro		2382
QY	455	TCAAAGCGCGCTGTATGTGGGCCATTCAAGCTCGCGCGGTCAATGCGGACTTGC		514
Db	2383	AlaProProProPro-----SerAlaGlnThrThrLeuProArgProAlaAlaProPro		2400
QY	515	AGCGCCCGCGCAACCGACGAGAATGCGGCCATGCGGGACCTGCGCCGACGAAAGCATG		574
Db	2401	AlaProProProProSerAlaGlnThr-----ThrLeuProArg-----ProAla		2415
QY	575	CCAGCGCGCCATCGGCAATTCGACCGCGGCGCTTCACCGCGCGCGCG		631
Db	2416	ProProProProSerAla-----ProAlaAlaThrProThrProProAlaProGluPro		2433
QY	632	CCACCGAAGATCATCGAGTGTGCGGCGCCCTGACGCGCATGCGGCATTAACGCA		691
Db	2434	AlaProSerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla		2453
QY	692	CCCAKATGCGCGCAAGGCGACGACATGTGCGCGCGCGTGGAGAACTTCGCGATCG		751
Db	2454	-ProAlaValAlaGAsAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla		2473
QY	752	GCGG-----CGAGCTGGAGCTGCGCGGTGTATCTCGACACACAG		793
Db	2473	grProLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla		2491
QY	794	TCATGCGCGACCGCAATTGCG		815
Db	2491	rValGluProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla		2511
QY	815	-----	-----	815

Db	2511	aseraLaserSerSerLeuLaLaSerThrProSerArgGluLaLaLaSerSerSerSer	2513
QY	816	-----CCGCTCGCGAGAGCTCCGC	838
Db	2531	rProLaagLYthrProSerAspGLyLaLaArgLYglYProProSerLeuLaLaGLyLaLa	2551
QY	839	TGATCGAGGGCCGATGCGCGCGAGACGTCGCTGGAGCGGTATCCCTACGTGGCG	898
Db	2551	aLaLeArgLaLProProdiLYaLaProGlySerGLyGLyLnaArgLaGlyLaSpGLyLyr	2571
QY	899	GCTCCACCATGCTCAACAGCAGGAGCGCTGCTGGCGCGGAGCAGCATACCTGGT	958
Db	2571	oLaLaLaSrpLaSerLeuArgLYaSpGLyProLaGLyLys--LeuProLaLaLeProLa	2599
QY	959	GCAAGCCTTCCCGAACTGAAGCGGGCGGCACTGGATGAATCGCGCGCGCGGCA	1016
Db	2591	LaLeGlnProLeuPProSerLeuLaValLeuArgThrLYaLaLaArgThrSerValGLyLeArg	2610
QY	1019	AATCAAGTACAGCATGTGATGCGGAGTGCAGCGCGCGCGGCGCATCTA-----CT	1066
Db	2611	AspHisSer-ArgArgLYuSerMetLaLeuLaProLaArgProValSerLaLProAr	2633
QY	1070	TCATGATGAGCAAGCCGAGCGTGCAGGCACTCCGTGGCTTCGGCCGAGCATATCGCT	1129
Db	2630	gGLnThrGLyLYeupProProLaLaLaGlnProGluArgProProGlyLYeGL	2650
QY	1130	CCGAGCGCT-----GCGCAGCAGCAGCGCCGATCCGGCTGTGGGCACTTC	1183
Db	2650	uProArgProProGlnLaGlnLaGlnLaProLaThrLaLProLaThrGLyPro--Pr	2663
QY	1184	CGCGGTGCTGGGGCATATGCGCGGACCTGGCGCTGTCC--GCTGAGAGCGGG	1240
Db	2669	oLa	2683
QY	1241	TATGGAAGATGACCGGCTGACCGCGCGCGCTCGCGCTGGCGCGGCGGCACTGC	1300
Db	2689	y-----ProValArgArgLaLaLaGLyPro-----ArgProProLa	2701
QY	1301	AGCGCGGGA-----CTCGCGCACTGTGTCTCG	1333
Db	2701	agLYnLaLeuArgLeuHisLaLeProProProGlnProGlnLeuGluGLyValLa	2721
QY	1334	ACCGGG--CCAGCGTGGCG--ATACCGCACCTTGACACCCATACAGAGCGCGCG	1389
Db	2721	lProLeuProLaSerProdiuThrProLaLaProLaGlnThrGlnProProArgSerPr	2741
QY	1390	GGCATCCATTCG-----	1402
Db	2741	oLeuLaLaLProThrSerLeuLaLaLProSerGluLeuArgProLaLaLaSerLa	2761
QY	1403	-----TGTACGTCAAGCGCGCGCGCTCGGCACAGCAGCGCG	1440
Db	2761	aLa	2781
QY	1441	TTCACCGCGAGATGCGCGCGCGCTGCTCGCAGCAGCGCGCG--CCTGAGCCGCG	1494
Db	2781	rLaLaProLaLaLProProProLaLaLProProProSerArgProLaLaLaLProProLa	2801
QY	1495	CGCCAGCCCTTCAATCCGCGGTGAAGGGGGCGT-----	1531
Db	2801	aProProProSerArgProLaLaLProThrProGluArgHisGLyLupProLaArgLeu	2821
QY	1532	-----GCGCGCCCTCCCAACCTG	1551
Db	2821	tPheThrProProProLeuProProGlnProSerGlnGlnArgProProGlnLaLaProTr	2841
QY	1552	GAGCAAAACGCTACATGGCCCTC-----CCTCGCTCGCAATACG	1593
Db	2841	p-----ThiTrpProGluProLaArgAspSerLaLaGLyProProLaLaLeuPheTh	2857
QY	1594	GCCCCACCGATATCGTGGGCAAGAAAGTATGGCGCGCGCTGCGCGAGCGCAG	1653













Db 526 ro-----ProProProGlyProProProProGlyProAlaProProG 540  
 QY 1728 GGGCCAGATCGCC 1741  
 Db 540 lYAlaArgProPro 544  
 RESULT 15  
 041972  
 ID 041972 PRELIMINARY; PRT; 706 AA.  
 AC 041972;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 GN Hypoetical 66.4 kDa protein.  
 OS GAMMAY.M10B.  
 OS murid herpesvirus 4.  
 OC Viruses; dsDNA viruses, no RNA stage: Herpesviridae;  
 OC Gammaherpesvirinae.  
 OX NCBI\_TaxID=33708;  
 RP (1)  
 RC SEQUENCE FROM N.A.  
 RA STRAIN-WDMS;  
 RA MEDLINE=97366649; PubMed-9223479;  
 RA Virgin H.W. IV, Latreille P., Wamsley P., Hallsworth K., Weck K.E.,  
 RA Dal Canto A.J., Speck S.H.;  
 RT "Complete sequence and genomic analysis of murine gammaherpesvirus  
 68."  
 RL J. Virol. 71:5894-5904(1997).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WDMS;  
 RA Latreille P., Wamsley P., Waterston R.H.;  
 RA Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U97553; AAB6421.1; .  
 KW Hypoetical protein.  
 SQ SEQUENCE 706 AA; 66413 MW; 9AD7A4AFDEB8748B CRC64;  
 Alignment Scores:  
 Pred. No.: 1,54e-07 Length: 706  
 Score: 299.00 Matches: 215  
 Percent Similarity: 33.43% Conservative: 20  
 Best Local Similarity: 30.58% Mismatches: 237  
 Query Match: 9.22% Indels: 231  
 DB: 12 Gaps: 50  
 US-10-009-782-1 (1-1758) x 041972 (1-706)  
 QY 1750 CGTAGCTCAGGCGCATCTGGCCCA-----GCTGGCCTTGACAGGCTGAGA---CCG 1700  
 Db 45 ArgGlySerGlyLeuProGlyProGlyCysGlySerProGlyLeuGlySerArgSerPro 64  
 QY 1699 CGATGCCGCTGGCCTCGACAGAGCTTGACAGGCTCATTTCCGCGCTTCGCTGGCGCG 1640  
 Db 65 ProAlaGlyTrpGlyLupProGly-----GlyArgGlyProGlyAlaArgGlu 80  
 QY 1639 GCAGGCGCGCCGCGCATCTCTCTCCACGATATCGGCTGGCGCGCGATTCGAGCGG 1580  
 Db 81 ArgAlaAlaArgProGlyAlaArgValProArg----- 91  
 QY 1579 AGGAGAGGCGCATATGAGCGTTGGCTCA-----GGTTGGAGAG---GGCGGACCGCG 1526  
 Db 92 -----ProGlyLupProValProProGlyGlyMetGlyGlyThreGlyLupGly 108  
 QY 1525 CCGCGTTCAGCGCGCATATGAGCGCTGGCGCGCGGCTCAGCGCGCGCGTGGCGAGCA 1466  
 Db 109 ProArgSerGlu-----GlyAlaGly-----CysProAla 118  
 QY 1465 GCGCGCGCGCATGCTGGCGCGGTGAACGCGTCTTT---GCCAGACGCGCGCGCGCTTGA 1409  
 Db 119 ArgGly-----AlaGly-----ProProAlaTrpGlyAlaGlyProProArg----- 132  
 QY 1408 CSTACACGGAATGATGCGCGCGCGCGCTCGGTAGGCTGTTCGAAGTGCGGTATCGG 1349

Db 133 -----ArgAspGlyGlyAlaAsp-----GlyAlaProGlyLupGlySerGly 148  
 QY 1348 CCACCGTGGCGCGGTCGAACACACCGAGTGGCGGAGGATCCGCGCGCGCGCGCGCG 1289  
 Db 149 LeuProGlyProGlyCysGlySerProGlyLeuGlySerArgSerProProAlaGlyTrp 168  
 QY 1288 GCCCGCGCGCG-----CGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1253  
 Db 169 GlyLupProGlyGlyArgGlyProGlyAlaArgLupArgAlaAlaArgProGlyValArg 188  
 QY 1252 -----TCATCTTCATACCGCGCG-----TCTCCACGCGGAGACAGCG 1217  
 Db 189 ValProArgProGlyLupProValProProGlyGlyMetGlyGlyThreGlyLupGly 208  
 QY 1216 CCAGTTCGCGCGCATAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1160  
 Db 209 Pro-----ArgSerGlyGlyAlaGlyLysProAlaAlaArgGlyAlaGlyProProAlaTrpGly 227  
 QY 1159 GCGGCG-----GCTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1130  
 Db 228 AlaGlyProProArgArgAspGlyGlyAlaArgGlyAlaAspGlyAlaProGlyLupGlySer 247  
 QY 1129 -----AGCCGATCA----- 1121  
 Db 248 GlyLeuProGlyProGlyCysGlySerProGlyLeuGlySerArgSerProProAlaGly 267  
 QY 1120 TGG-----TCGCGCGGAGCGCGAGATGCGCTGCGCGCGCGCGCGCGCGCGCG 1070  
 Db 268 TrpGlyLupProGlyGlyArgGlyProGlyAlaArgLupArgAlaAlaArgPro----- 285  
 QY 1069 AGTAGATGCT 1010  
 Db 286 -----GlyValArgValProArg----- 291  
 QY 1009 CGCGCGGACTTCATCCAGTTCGC----- 986  
 Db 292 -----ProGlyLupProValProProGlyGlyMetGlyGlyThreGly 306  
 QY 985 -----GCCGCTAGTTCGGGAGAGGCTTGACACAGATGATGCTGCGCGCGCA 932  
 Db 307 ThrGlyProArgSer-----GlyGlyAlaGlyLys-----Pro 317  
 QY 931 GCAGCAGCGCGTCTGCTTACGATGCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 872  
 Db 318 AlaArgGlyAlaGlyProProAlaTrpGlyAlaGlyProProArgArgAspGlyAla 337  
 QY 871 AGAGTCTCTGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 812  
 Db 338 ArgGlyAspGlyAlaProGlyLup-----GlySer 347  
 QY 811 AATGGCGTGGCGCGCATGACCTGT-----GGTGGAGATCACACCGCGCA 767  
 Db 348 GlyLeuProGlyPro-----GlyCysGlySerProGlyLeuGlySerArgSerProProAla 366  
 QY 766 -----GTCACACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 716  
 Db 367 GlyTrpGlyLupProGly-----GlyArgGlyProGlyAlaArgLupArgAla 382  
 QY 715 GTCGCGCTTCGTCGCGCATGCTGGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 656  
 Db 383 AlaArgProGly-----ValArgVal-----ProArgProGlyLupProVal 396  
 QY 655 ACACCTCATGATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 596  
 Db 396 LProProGlyGlyMetGlyGlyThreGlyLupGlyProArgSerGlyGlyAlaGly 416  
 QY 595 AATGCGCATGCG 536  
 Db 416 sProAlaArgGly-----AlaGlyProProAlaTrpGlyAlaGlyProProArgArgAsp--- 434  
 QY 535 CCTGCTGCTGCG 476

Db 435 -----GlyGlyAsnArgGlyAspGlyAlaProGluArgGly--SerGlyLeuProGly 451  
QY 475 CCA-----CCATACAGCGCGCGTTCAGCGCGCGCG-----TGC-----CCC 437  
Db 452 ProGlyCysGlySerProGlyLeuGlySerArgSerProAlaGlyTrpGlyGluPro 471  
QY 436 GCAAGCGGT---CCAGTAGTCGG---CGAAGGCTCGAAGCGTAGAAGCCGCTTCGT 383  
Db 472 GlyGlyArgGlyProGlyAlaArgGluArgAlaAlaArgProGlyAlaArgValProArg 491  
QY 382 CCAGCAGTCCA-----GGGGGGCGGGGGTTCGCGCGCGCGCG 341  
Db 492 ProGlyGluProValProProGlyGlyMetGlyGlyThrGlyGly---ThrGlyProArg 510  
QY 340 GCGCCAGGCTGATGCCGCAATTGCCG-----TGACCACCGTGGTAGCGCCT 293  
Db 511 SerGluGly-----AlaGlyCysProAlaArgGlyAlaGlyProPro----- 524  
QY 292 GCGAGATCTGGCGCGCATGTCCGACGCGCTGACGAGGTAGTGTGTGCTGT----- 242  
Db 525 -----AlaTrpGly-----AlaGlyProProArgArgAspGlyGly 536  
QY 241 -----GGGTGTGCGAGT 230  
Db 537 AsnArgGlyAspGlyAlaProGluArgGlySerGlyLeuProGlyProGlyCysGlySer 556  
QY 229 CGATGAAGCGCGCGCGCA---CCACCAAGCGCCGACACGTGACCGCGGTGGCGCGCG 173  
Db 557 ProGlyLeuGlySerArgSerProAlaGlyTrpGlyGluProGlyGlyArgGlyPro 576  
QY 172 CGTCCGACAGATGCGCGATGCGCGATGCGCGGTGCGCGCGCGCGCGCGCGCC 113  
Db 577 GlyAlaArgGluArgAlaAlaArgProGlyValArgValProAlaArgProGlyGluProVal 596  
QY 112 GCCCGGGGTGTGTCGCGCGATGAGGGTGCAGCGCGCGCGCGCGCGCGCGT----- 62  
Db 597 ProProGlyGlyMetGlyGlyThrGlyGlyThrGlyProArgSerGluGlyAlaGlyCys 616  
QY 61 -----CGAAGGCTCGGATCGGATGGGACATGGAATCTCTCCGCGCATCAAGT 8  
Db 617 ProAlaArgGlyAlaGlyPro-----ProAlaTrpGlyAlaGlyProProArgArgAsp 634  
QY 7 GGA 5  
Db 635 Gly 635

Search completed: May 11, 2003, 12:15:53  
Job time : 113 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 01:33:19 ; Search time 82 Seconds

(Without alignments)

6574.851 Million cell updates/sec

Title: US-10-009-782-1

Sequence: 1758

Sequence: 1 gaattcaccatgacgcgcga.....ccctgagctacgagaagctt 1758

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/PCRTS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description
1	94.2	5.4	1	Sequence 1, Appl1
2	81.8	4.7	1248	Sequence 7, Appl1
3	81.8	4.7	13613	Sequence 3, Appl1
4	77.4	4.4	23673	Sequence 1, Appl1
5	76.2	4.3	5970	Sequence 21, Appl1
6	75.6	4.3	43280	Sequence 1, Appl1
7	75	4.3	1320	Sequence 8, Appl1
8	75	4.3	1320	Sequence 8, Appl1
9	75	4.3	1320	Sequence 8, Appl1
10	75	4.3	1620	Sequence 10, Appl1
11	75	4.3	1620	Sequence 10, Appl1
12	73.8	4.2	15872	Sequence 11, Appl1
13	73.4	4.2	2167	Sequence 9, Appl1
14	73.4	4.2	2167	Sequence 9, Appl1
15	73.4	4.2	2668	Sequence 11, Appl1
16	73.4	4.2	2668	Sequence 11, Appl1
17	73.4	4.2	4257	Sequence 11, Appl1
18	73.4	4.2	4257	Sequence 11, Appl1
19	73.4	4.2	4257	Sequence 11, Appl1
20	72.8	4.1	11604	Sequence 1, Appl1
21	72.8	4.1	15079	Sequence 1, Appl1
22	72.2	4.1	13842	Sequence 13, Appl1
23	72.2	4.1	36778	Sequence 30, Appl1
24	72.2	4.1	38506	Sequence 5, Appl1
25	71.2	4.1	1833	Sequence 19, Appl1
26	71.2	4.1	1833	Sequence 6, Appl1
27	71.2	4.1	1833	Sequence 6, Appl1

28	71.2	4.1	2634	US-08-941-936-1	Sequence 1, Appl1
29	70	4.0	4403765	US-09-103-840A-2	Sequence 2, Appl1
30	69.6	4.0	11220	US-09-105-537-32	Sequence 32, Appl1
31	68.8	3.9	5392	US-08-403-852D-1	Sequence 1, Appl1
32	68.8	3.9	5392	US-08-403-852D-1	Sequence 1, Appl1
33	68.8	3.9	5392	US-08-403-852D-1	Sequence 1, Appl1
34	68.8	3.9	5392	US-08-403-852D-1	Sequence 1, Appl1
35	68.6	3.9	44377	US-08-804-198-1	Sequence 7, Appl1
36	68.2	3.9	2220	US-08-804-198-1	Sequence 1, Appl1
37	68.2	3.9	4496	US-08-765-907A-14	Sequence 14, Appl1
38	68	3.9	2064	US-08-343-928-1	Sequence 6, Appl1
39	68	3.9	12588	US-08-387-942C-1	Sequence 1, Appl1
40	67.8	3.9	1014	US-09-105-537-13	Sequence 1, Appl1
41	67.4	3.8	30001	US-08-125-468-1	Sequence 13, Appl1
42	67.4	3.8	30001	US-08-474-933-1	Sequence 1, Appl1
43	67.2	3.8	1208	US-08-403-852D-4	Sequence 4, Appl1
44	67.2	3.8	1208	US-08-510-646B-4	Sequence 4, Appl1
45	67.2	3.8	1208	US-09-231-818-4	Sequence 4, Appl1

#### ALIGNMENTS

RESULT 1  
US-07-945-283-1/C  
Sequence 1, Application US/07945283

Patent No. 5352596

GENERAL INFORMATION:

APPLICANT: Cheung, Andrew K.

TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants

TITLE OF INVENTION: Involving The EPO and LIT Genes

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Curtis P. Ribando

STREET: 1815 No. 5352596th University Street

CITY: Peoria

STATE: IL

COUNTRY: USA

ZIP: 61604

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/945,283

FILING DATE: 19920911

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Ribando, Curtis P

REGISTRATION NUMBER: 27976

TELEPHONE: 309-685-4011 ext.513

TELEFAX: 309-685-4128

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 8438 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORGANISM: Pseudorabies virus

FEATURE:

NAME/KEY: CDS

LOCATION: 622..6495

FEATURE:

NAME/KEY: variation

LOCATION: replace(1099, "g")

Query Match	5.48;	Score 94.2;	DB 1;	Length 8438;
Best Local Similarity	44.5%;	Pred. NO. 8.e-10;		
Matches 608; Conservative	0;	Mismatches 738;	Indels	21; Gaps 5

QY	868	GTCTCGCTGAGACGGATGCCACAGC-----TGGCCGGCTTCACACATGCTCAAGACAGACCG	924
Dd	4767	CGGATCCGGCGCCGCTTGGCCGGCCGGACGGCCCGGCGCGTGGCCCCCGAGCTGGGGAGC	4701
QY	925	GTGTGCTGTGGCCGAGGACCATCATCTACCTGTGTGAAGCCCTTCCCGCAATCGAAGGGG	984
Dd	4707	GCTTGGTCTCTGGCTCGCCGGCGCGTCTCTGAGGGCCCTGTGCTTGTCCGCTTCGGGTG	4641
QY	985	CGCGACCTGGATGAAGTGGCGGGCCGGAGCGCGGCAATCCAAATGACGAGTGTGCCGAG	1044
Dd	4647	CCCCGGCCCCCGCCGGCCCCCGGCGCCCTGGGGCCCGCTGTGCTTCGAGGAGTGAACGGCC	4588
QY	1045	CTGCACCGCGCCGGCGCCATCTACTTATGATGAGACGACCCGAGTGCAGCGGCACTCTG	1104
Dd	4587	GGCGTGTCTCGCGCTTCCGGAGCGGATCCCGGGGGCCGGCCCGGCGGACGACGAGCC	4528
QY	1105	GGCTTCGGGCCCGACATGATGTGGTCTCGACGGCTTGCCTGCACGACGAGCGCCCGATCCG	1164
Dd	4527	CGGACCTTGTTCGCCCTGTGTGGCCCGGACCGTGGCCCGC-----TGTGTGGTACAG	4475
QY	1165	CGCCTGTGGGACACTTTCGCCGGGGTCTGGGCGACTATGCGCGGACACTTGGGCTTTC	1224
Dd	4474	CGTGGACGGGGCCCGGCCCGCGC---AGCCCGCTGGACCTACGCCCCGGCCCTTTC	4420
QY	1225	CGGCTGAGAGACGGGTATGTGAAGATGACCGGCTTACCGCGCGCGCTTCTGGGCTTGGC	1284
Dd	4419	GCCCCGGCCACGTGGGCGGGGGCCGGCGCTGCGAGAGCGCGCGCGCGGCGGGCGCGG	4366
QY	1285	GGGCGCGGGGACGTTCGAGCGCGGGTACTTTCGCGGACCTGTGTGTGTGACCCGGCCAGC	1344
Dd	4359	GAGCGCCGCGCCCGGCGCTGCCCGCGCTCTGGCCCCGAGCAACCCGGCTGTGTCTCCACG	4300
QY	1345	GTGGCGGATGCCCACTTTCGAACACCTTACCGAGCGCGCGCGG	1391
Dd	4299	CGGCGCGCGGGCGGGCGGGCGCCCTTCGCGCTCTCCCGCTCCGGTCCGG	4253

	Query Match	4.7%	Score 81.8;	DB 4;	Length 1248;
	Best Local Similarity	44.2%	Pred. 1.6e-07;		
	Matches 439;	Conservative 0;	Mismatches 542;	Indels 12;	Gaps 2
QY	32	CCATGTCCTCCAAATCCCATTTCCGAGCCCTTGACACTCTGCTGTCCGGGGGACACCTTCATCG	91		
Db	215	CCACCTGCAAGCCACGAGCCGGGGCTTCACACTCTTCGGCGACGCCCGCCGCTCACCGGGG	274		
QY	92	ACGGCAGCAACACCCCGGGGGCGCGCCGACCTGGGGGTGGCGCGACACGACCTGACCGG	151		
Db	275	AAGGATATCATCCCTGCTGATGAGCTTGGCGCCACCCGACGACGACTGGCTGGATCGGCC	334		
QY	152	CCATCGGCGATCTGTGGACGCGCGCGGCGACACCGGGGTGAGAGTGTGGGCGCTGTGG	211		
Db	335	TCACCCGGGCTTGGCGCGACATCGACCCGAGACACCGCGACACTGACCCGAGCAGGGTG	394		







## RESULT 5

US-09-320-878-21/c  
 ; Sequence 21, Application US/09320878A  
 ; Patent No. 6117659  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ASHLEY, Gary  
 ; APPLICANT: BETLACH, Melaine C.  
 ; APPLICANT: BETLACH, Mary C.  
 ; APPLICANT: MCDANIEL, Robert  
 ; APPLICANT: TANG, Li  
 ; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
 ; FILE REFERENCE: 300622002120  
 ; CURRENT APPLICATION NUMBER: US/09/320, 878A  
 ; CURRENT FILING DATE: 1999-05-27  
 ; EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
 ; EARLIER FILING DATE: 1998-08-28  
 ; EARLIER APPLICATION NUMBER: CIP OF 09/073, 538  
 ; EARLIER FILING DATE: 1998-05-06  
 ; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
 ; EARLIER FILING DATE: 1997-04-30  
 ; EARLIER APPLICATION NUMBER: 60/119,139  
 ; EARLIER FILING DATE: 1999-02-08  
 ; EARLIER APPLICATION NUMBER: 60/100,880  
 ; EARLIER FILING DATE: 1998-09-22  
 ; EARLIER APPLICATION NUMBER: 60/087,080  
 ; EARLIER FILING DATE: 1998-05-28  
 ; NUMBER OF SEQ ID NOS: 34  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 21  
 ; LENGTH: 5970  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-320-878-21

Query Match 4.3%; Score 76.2; DB 3; Length 5970;  
 Best Local Similarity 44.7%; Pred. No. 2,2e-06;  
 Matches 334; Conservative 1; Mismatches 409; Indels 3; Gaps 1;

QY 32 CCATGTCCTCCATGCGATTCGACGCTTCTGACCTGCTGCGGGGACACCTTCATCG 91  
 DB 781 CCACTGGAACGCCACGCGCGGCTCCAGCTCTCGCGACGCGCGGCTCCACGCGG 722  
 QY 92 ACGGAGCAACACCGCGGGGGGCGCGGCGGCTGCGGCGCTGCGCGGCGGCGGCGG 151  
 DB 721 AAGTATATATGCGCTGATGACGTTGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 662  
 QY 152 CCATGCGGCGATCTGCGGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 211  
 DB 661 TCACCGCGGCTCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 602  
 QY 212 TCGGCGCGGCGCTTCATGACGCTGCGACACCGGCGGCGGCGGCGGCGGCGGCGG 271  
 DB 601 CCGCGCGGCTGACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 542  
 QY 272 ACATACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 331  
 DB 541 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 482  
 QY 332 GCGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 391  
 DB 481 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 422  
 QY 392 GCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 451  
 DB 421 AGGTCCTGAGCTTC---CAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 365  
 QY 452 CCGTAAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 511  
 DB 364 TCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 305  
 QY 512 TGCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571

DB 304 TGCCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 245  
 QY 572 TGGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631  
 DB 244 TGGCGCTACCTCTCTGAGCGGCTTCCGAGCTCATCGACCGGCGGCGGCGGCGG 185  
 QY 632 CCACGGAAGATATCATGAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 691  
 DB 184 CCGGCTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 125  
 QY 692 CCACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 751  
 DB 124 ACGGCTCAACCAACCAACGATGATGATGATGATGATGATGATGATGATGATGATG 65  
 QY 752 GCGCGAGCTGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 778  
 DB 64 ACCGCGACCTGCTCATGAGGTCTCTA 38

## RESULT 6

US-08-804-227C-1  
 ; Sequence 1, Application US/08804227C  
 ; Patent No. 5876991  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dehoff, Bradley S.  
 ; APPLICANT: Kustoss, Stuart A.  
 ; APPLICANT: Rostock, Paul R., Jr.  
 ; APPLICANT: Sulton, Kimberly L.  
 ; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
 ; NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: THOMAS G. PLANT 1501  
 STREET: LILLY CORPORATE CENTER  
 CITY: INDIANAPOLIS  
 STATE: IN  
 COUNTRY: USA

ZIP: 46285  
 COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: ASCII(DOS) Text only

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/804,227C  
 FILING DATE: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 43280 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 816..14234

FEATURE:

NAME/KEY: CDS

LOCATION: 14351..19945

FEATURE:

NAME/KEY: CDS

LOCATION: 20010..31199

FEATURE:

NAME/KEY: CDS

LOCATION: 31232..36067

FEATURE:

NAME/KEY: CDS

LOCATION: 31232..36067

35 TGTCCCAATCCGATTCTCCAGACCCCTTGACCTGTGCTCGCGGGCGGACACCTCATGACG 94  
 Db 28810 TGGACCCCGGGCGGGCCCGGCGGAGAGACCTGTACCGCCCGGACACACCGGCTGCG 28869  
 QY 95 GCAGCAACACCCCGGGGGCGCGCGCCGAGACTGTGGGCGTGGCGGCGACCGCATTCGCCGCA 154  
 Db 28870 GGGTCACCCCGCGGGGGGGAGCGCTGGAGAACTGTCTGCTCGCGCCCGCGCCGACGCG 28929  
 QY 155 TCGGCGATTTGTTCGAGCGCGCGCGCGCAACCCGGGTTCAGCTGTGCGGCGCTGTGATG 214  
 Db 28930 AGGAAACCATGCGACCGGGCGCAAGATACGATGCGCGTCCGCGCGGCGGCGTGAACCTTC 28989  
 QY 215 CGCGCGGCTTACATCGACTCGACACCCACAGCAGCAACTACTCTGTCAAGGCGTCGCGACA 274  
 Db 28990 GCGACCGCTGTATCGCCTTCGCGCATGTACCCGGGCAA-----GGGAACATAGGCGCG 29043  
 QY 275 TGACGCCCAAGATCTCGCGAGGCGGTACACAGGTGTACAGGGGCAATTGGCGCATCAGC 334  
 Db 29044 AGGAGACCGGGCGTGTCTGTTCGAGACCGCCCGCATGTCAACGCGCTTCGCGCGGAGAC 29103  
 QY 335 TGGCGCGCTGGCGCGACGCCAACCGCGCCCGCCCTGTGACTGTGTGACGAAAGCGGCT 394  
 Db 29104 GCGTCTCGCATGTGGAACGGCGGCTTCGGGCGCCCTGTGTGGCGGACACCGCATAG 29163  
 QY 395 CTTACCGTTTGAAGGGCTTCGCGCACTACACTGTGACAGCTTGGCGGCCACGCGCGGCGG 454  
 Db 29164 TGGCCCCGATCCCCACAGCGGTGTGTACGCGCGAGAGCGGCTCTCGTCCCGCGTCTC 29223  
 QY 455 TCACGCGCGCTGTATGTGTGGGCCATTTCAGCTGCGGGCGGCGGTCTATGGCGGACTTGC 514  
 Db 29224 TCACCTCTTACTACGCGCTGACCGCGGCTGTGGCGCGCCGCGCGCGGACGACAGACCTTCTCG 29283  
 QY 515 AGCGGCGCGCCACCGACAGAGAAATCGCGCCATCGGGACCTGGCGGAGAACCATAG 574  
 Db 29284 TCCACGCGCGCGCGCGCGGTGTGCGCATGTGGACCTTCAACTCGCGCGCACCTCGGCG 29343  
 QY 575 CCAGCGGGCGCATCGGCAATTTTCAGCCGCGGCTTCTACCGCGCGCGCGCGCGCCACCA 634  
 Db 29344 TGGAGGTATAGCCACCGCGGACACCGGCAATAGGAGCGCTTCAGAAAGCACCGCATTC 29403  
 QY 635 CCGAAGAGATCATGAGGTGTGCCGCGGCTGAGCGCGCATGGGCGCATTCAGCCACCG 694  
 Db 29404 CCGAGGACCGCATCGCGGACGCTCCGCAACCGT---GACTTCGCGGAGCGCTTCTGTGCC 29460  
 QY 695 ACATGCGGAGAGAAAGGCGAGCATATCGTGGCGCGGTGAGGAAACCTTCGCACTCGGC 754  
 Db 29461 GAGAGGGGCGCGGGGTGTGCACATCGTGTGAATCTCTGCGCGCGGAGTGTGTGACG 29520  
 QY 755 GCGAGCTGAGAGTCCGCGGTGTATCTCGACACCAAGGTCAATGGGCGACCCCAATTGCG 814  
 Db 29521 CCTACTGCGGCTCTCTCCGCGCGCGGCGCACTTCTTGGAATCGGCAAGCGGACGATTC 29580  
 QY 815 GCGCGCTCGCGAGACGCTCGGCTGATTCAGAGCGCGCATGTGGCGCGCAAGAGCTGCGC 874  
 Db 29581 GCGACCCCCCGGAGATGCGCGCGCGCCCATCCGGGCAACGACTACCGGCGCTTGTGACCTGG 29640  
 QY 875 TGGACCGGTATCCTTACGTGGCGGCTTCACCATGTCTCAAGAGAGAACCGGCTGTCTGG 934  
 Db 29641 TGCAGGCGGATCCGACACGATGCGGGAGATGCTCGGGAACTGTGGAACCTTGTGGCG 29700  
 QY 935 CCGGAGCACCATCATCACTGTGTGAAGGCCCTTCCCGAACTGAGCGGGCGGACCTGG 994  
 Db 29701 CCGGAGGCGCTGCGCGCGCTCCGCTCACCGGCTACGCGCATACGACGCCCGGACG---- 29756  
 QY 995 ATGAAGTCGCGGCGGACGCGGCAATTCAGATACGAGCTGTGCCGAGCTCACCGG 1054

D	29757	----	GCCTTGGGACACCCCTCAGCCAGGCCCCGGCACACCGGCAAGCTCGTGTGACGGTGC	29811
Q	1055	CCGGGCGCATCTACTTTCATGATGAGGAACCCGACCTGCAAGCGCATCTGCGGTTGCGGC	1114	
D	29812	CTGCCGGATTTCGACACACCACCGACAGGTGCTCTCTCAACGGGGGACAGGGGACAGCTGCGCC	29871	
Q	1115	GCACCATGATTCGGCTTCGACGGCTTCGCCACAGCAGAGCGCCCGCATCCGCGCCTGTGGG	1174	
D	29872	ACACACTGCGCCCGGCATCTGGTGTCAACCCGACCGGCGTACGCGACCTGCTGTGCGCGGCC	29931	
Q	1175	GCACCTTCCCGGGGGTGGCTGGGGGCACTATAGCGGGGACCTGGGCGCTGTTCGCCGTGGAGA	1234	
D	29932	GCACCGCGCGCGCGCGCCGACAGGCGCGTCCGCGGAATATGTGTACCTTGGGCGAGTTGGCGC	29993	
Q	1235	CGCGCGTATGGAAGATGACCGGCGCTGACCGCGCGCGCTTGCGCTGGCGC--GGCGCG	1291	
D	29992	CCGAGGTCGGGGTCCGGCGCTCGACAGGGGGCGGACCGCGAGGGGCTCACCGAACTTCCTCG	30051	
Q	1292	GGCACTGCAAGCGCCGGGTACTTCGCCGACCTGTGTGTTCGACCCGGGACAGGTGGCCG	1351	
D	30052	CCGGAAATCCCGCTGCAGACACCGCTCGGCGCGCTGTGCACGCGCGGGACACTTCGAGC	30111	
Q	1352	ATACCGGACACCTTGCAACACCCCTACCGAGCGCGCGCGCGCATTCCTCGGTACGTCA	1411	
D	30112	ACGGCACCATATCCCTCACTGACCGCGCGGAACATGCAACAGTGTGGCGCCCAAGGCGCG	30171	
Q	1412	ACGGCGCGCGGATCTGGCGAAGAGCAGGGGTTACCGCGCAGCATCCGCGCGCGGTGCG	1471	
D	30172	ACGGCGGTCTCACTCACTGACAGAGGTGACCCCGAGAGCGGCACTCTCGGCGTTCTGCTCTC-	30230	
Q	1472	CACCGACGGCGCGCTGAGACCCGGCGCCAGCCCTTACAAATCCGGCGTGAAGGGGCGCGGT	1531	
D	30231	--TACTGTCTCTCTCGCGGTGCTGTGCGACCCCGGCAAGGGGCGCTTACGCGCGGGCCA	30288	
Q	1532	GCGCGCGCGCTCCCAACCTCTGAGCGAAACGCGTCAATGGGCGCTTCCCTCCGCTGGCAATA	1591	
D	30289	ACGGCTTCTCTGAGAGCGGCTTGGCCCGGATACCGCAAGGCGCTTGGGCTGTGCGCGCTTTCGC	30348	
Q	1592	CGGCGCCACCCGATATCGTGGGCAAGGAAGTATGGGCGCGCGCTGTGCGCGCCAGGCA	1651	
D	30349	TGGCTCTGGGACTGTGGGGGCGCAACAGCCGCAATGGGGGCGCACTCGAACATGTCGGGCA	30408	
Q	1652	AGGCGCGGAAATATACCTCTGCAACACTGTTCGACAGGCCAGCGCATGCGGGTCTCGACC	1711	
D	30409	TGCACCGGCGCCTGTAACCGGAGCGGCAATCATGTCGCTCACGAGCGCGAGGGGCTTCGCC	30468	
Q	1712	TGTCCAGGCGCGAGCTGGGC	1731	
D	30469	TGTTCCAGCGCGCACAGGAC	30488	

RESULT 7  
US-08-461-775-8  
Sequence 8, Application US/08461775 .  
Patent No. 5858773  
GENERAL INFORMATION:  
APPLICANT: MACODIER, Philippe  
APPLICANT: GUGLIERMI, Gerard  
TITLE OF INVENTION: REGULATORY NOCETIDE SEQUENCE OF THE  
TITLE OF INVENTION: INITIATION OF TRANSCRIPTION  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,775
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/050,313
FILING DATE: 10-MAY-1993
APPLICATION NUMBER: FR 9011186
FILING DATE: 10-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 010830-035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-2021
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1320
US-08-461-775-8

Query Match      4.38; Score 75; DB 2; Length 1320;
Best Local Similarity 46.9%; Pred. No. 3.2e-06;
Matches 234; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 297 CGTCCACAGGTGTGTCACGGGCAATTTGGGCGATCAGCTGGCGCCGCGTGGCGGACGCCAA 356
DB 90 CGGCCCCAAGGGCGCGCAACGTCGATCGACAGAAAGTTGCGCGCCCGACCATCACCAG 149
QY 357 CCGGCCCCCGCTGTGACCTGTGAGAGAGCGGCTCTTACCGTTTCGAGCGCTTGC 416
DB 150 CGAGCGGTGCACCATCGCCGTTGAGTGCAGTGCAGAGAGAGAGAGAGAGAGAGAGAG 209
QY 417 CGACTACCTGGAGCGCTTGGGGGCGACCGCGCGGCGCTCAACCGCGCTGTATGTGG 476
DB 210 CGAGCTGTCAAGAGAGTGGCGACCAAGACCAAGACATCGCGGGGTACGGACACCAAC 269
QY 477 CCATTCAACGCTGCGCGCGCGGTCATGCGGACTTGCACCGCGCGCGCGCGCGCGCG 536
DB 270 CGGAGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 329
QY 537 AATCGCGGCGCATCGGAGACCTTGGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 596
DB 330 CTCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 389
QY 597 GACCGGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656
DB 390 GCTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 449
QY 657 CCGGCGCGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 716
DB 450 CGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 509
QY 717 CATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 776
DB 510 CGGTCGATCAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 569
QY 777 GATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 795
DB 570 CATGCGCTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588

```

RESULT 8  
 US-09-031-606-8  
 ; Sequence 8, Application US/09031606  
 ; Patent No. 6153404

```

GENERAL INFORMATION:
APPLICANT: MAZODIER, Philippe
APPLICANT: GUGLIENT, Gerard
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
TITLE OF INVENTION: INITIATION OF TRANSCRIPTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: Burns, Doane, Swecker & Mathis
STREET: George Mason Bldg., Washington & Prince Sts.
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031,606
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/050,313
FILING DATE: 10-MAY-1993
APPLICATION NUMBER: FR 9011186
FILING DATE: 10-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 010830-035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-2021
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1320
US-09-031-606-8

Query Match      4.38; Score 75; DB 3; Length 1320;
Best Local Similarity 46.9%; Pred. No. 3.2e-06;
Matches 234; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 297 CGTCCACAGGTGTGTCACGGGCAATTTGGGCGATCAGCTGGCGCCGCGTGGCGGACGCCAA 356
DB 90 CGGCCCCAAGGGCGCGCAACGTCGATCGACAGAAAGTTGCGCGCCCGACCATCACCAG 149
QY 357 CCGGCCCCCGCTGTGACCTGTGAGAGAGCGGCTCTTACCGTTTCGAGCGCTTGC 416
DB 150 CGAGCGGTGCACCATCGCCGTTGAGTGCAGTGCAGAGAGAGAGAGAGAGAGAGAGAG 209
QY 417 CGACTACCTGGAGCGCTTGGGGGCGACCGCGCGGCGCTCAACCGCGCTGTATGTGG 476
DB 210 CGAGCTGTCAAGAGAGTGGCGACCAAGACCAAGACATCGCGGGGTACGGACACCAAC 269
QY 477 CCATTCAACGCTGCGCGCGCGGTCATGCGGACTTGCACCGCGCGCGCGCGCGCGCG 536
DB 270 CGGAGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 329
QY 537 AATCGCGGCGCATCGGAGACCTTGGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 596
DB 330 CTCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 389
QY 597 GACCGGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656
DB 390 GCTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 449

```





TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1620  
US-09-031-606-10

Query Match 4.38; Score 75; DB 3; Length 1620;  
Best Local Similarity 46.98; Pred. No. 3.3e-06;  
Matches 234; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY CGTACACAGGCTGCTACAGGCAATGCGGATCAGCTGGCGCCGCTGGCGCACGCCAA 356  
Db CGGCCCCAAGGCGCCGACAGCTGTCATGACAAAGATTCCGCGCCCGACATCACCAA 149  
QY CCGCGCCGCGCCCGCTGACCTGAGACGAGCGGCTTACCGCTTACGCGCTTGGC 416  
Db CGACGGGCTACCATCGCCGCTGAGTGAAGTGCAGACACCCGTAACGAACTTGGGCGC 209  
QY CGACTACCTGAGACCGCTTGGCGGCGCACCGCGCGCTCAACCGCGCTGATGTTGG 476  
Db CCACTCTGTACAGAGGCTGGCGACCAAGACCAAGACATCGCGGTTAGCGCACACAC 269  
QY CCATTCAAGCTGGCGCGCGCTGATCGCGGATTCGAGCGCGCGCGCACACGAGGA 536  
Db CGGACCGCTGCTGCGCCAGAGGCTGCTGCGCGGCGCTGCGCAACGCTGCGCGCGC 329  
QY AATCGCGGCTATCGGCGGCTGCGCGGAGAACCATGCGCGCGCGCGCATCGCATTTT 596  
Db CTCGCGCGCGCGCTGAGAAAGGCGATCGACGCGCGCGCTGCGCGCGCTGCGCGACT 389  
QY GACCGGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 656  
Db GCTGACACCGCTG 449  
QY CCGCGCGCTGAGCTG 716  
Db CGGCGCGGACAGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 509  
QY CATCGTGGCGCGCTGAGAGAACTTCCGATCGCGCGCGCGCGCGCGCGCGCGCTG 776  
Db CGGTGATCATCAGCTGAGAGAGTCAACACCTTGTGTGACTGATGATTCACCGAGG 569  
QY GATCTGCGCACACAGCTC 795  
Db 570 CATGCGCTTGCACAGGCG 588

## RESULT 11

US-08-458-568A-11/C  
Sequence 11, Application US/08458568A  
Patent No. 5821339  
GENERAL INFORMATION:  
APPLICANT: Schaffer, Priscilla A.  
APPLICANT: Yeh, Lily  
TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339r1s  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/458,568A  
FILING DATE: 02-JUNE-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/065,146  
FILING DATE: 05-MAY-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Leary Ph.D., Kathryn R.  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: DECI-0029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Herpes simplex virus  
STRAIN: Herpes Simplex Virus Type 1  
US-08-458-568A-11

Query Match 4.28; Score 74.4; DB 1; Length 12001;  
Best Local Similarity 42.65; Pred. No. 5.2e-06;  
Matches 581; Conservative 0; Mismatches 766; Indels 17; Gaps 3;

QY 189 GGTGACGTGTCGGGCGCTGCGGCGCGCGCGCGCTTATGATCGATCGACACCAAGCA 248  
Db 5125 GGTGACGTGTCGGGCGCTGCGGCGCGCGCGCGCTTATGATCGATCGACACCAAGCA 5066  
QY 249 CAACTACCTGTCAGCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 308  
Db 5065 CAGCAGCTCTTCAATCAACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5006  
QY 309 GGTGACGTGTCGGGCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368  
Db 5005 GGTGACGTGTCGGGCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4946  
QY 369 CTGACGCTGTCGAGACAGAGCGCGCTTACCGTTTGAACGCTTGGCGCATCTGGA 428  
Db 4945 CCAATACGACG 4886  
QY 429 CCGGTTGCGGCG 488  
Db 4885 CCGTTGGCG 4826  
QY 489 GCG 548  
Db 4825 AGATGACGAGAGGCG 4766  
QY 549 GCGGACCTGCG 603  
Db 4765 CCGGTTACG 4706  
QY 604 GCGTTTACCG 663  
Db 4705 CCGGCGGCG 4646  
QY 664 CTGAGCG 723  
Db 4645 CCG 4586  
QY 724 GCG 783  
Db 4585 GAGCG 4526  
QY 784 CACACAGGTGATGGCG 843

Db 4525 CCGCCCGCCAGACCCCGCCGCGGAGGAGACCCCGCGGCCCCCGCTTCCCGCCGCGCGC 4466  
QY 844 GAGCCCGCATGAGCGCCGAGAGAGCTCTGCTGAGAGCGGTATCCCTATGCGCGGCTTC 903  
Db 4465 CACCGCCAGCGCCCGCTCGCGCGCTGCTGCGAGCTGCGGCTTCGTCGCGGAGCGC 4406  
QY 904 ACATGCTCAAGAGAGAGCGGCTGTGTGCTGCGGAGCGAGCATCATCATCTGTCAG 963  
Db 4405 CTGTGCTCATG-----CGCTTCCGCGGAGACCTGCGCGCTGCGCGGAGAGGCC 4352  
QY 964 CCGTCCCGGAGTACGAGCGGCGCGAGCTGATGATGATGCGGCGCGAGCGCGCAATTC 1023  
Db 4351 GCGCTGAGCGCGCTGCGCGCGCTGAGCTGTGCTGCGCGCGCGCGCGCGCGCGCTCG 4292  
QY 1024 AAGTACGAGCTGTGCGCGAGCTGACAGCGCGCGCGCATCTACTTATGATGAGAA 1083  
Db 4291 CCGGAGCGCGCGCTGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTGACAC 4232  
QY 1084 CCGAGCTGAGAGCGATCTGCGCTGCGCGCGCGCGAGCATGATCGCGCTCGAGCGCTGCG 1143  
Db 4231 CAGAGCTGCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 4172  
QY 1144 CAGAGAGAGCGCGCGCATCGCGCGCTGTGCGGAGCACTTCCGCGCGCTGCGGCGACTAT 1203  
Db 4171 CCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4112  
QY 1204 CCGCGCGAGACCTGCGCGCTGTCCGCTGCGGAGAGCGCGGTATGAGATGACCGCGCTGAC 1263  
Db 4111 CCGCGCGCGCGCGCGAGCG 4052  
QY 1264 CCGCGCGCTTCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1323  
Db 4051 GCGCGCGCGCTGCGAGCG 3992  
QY 1324 GTGTGTGTCACCGCGCGAGAGTGTGCGCGATAC-----CGCGCATTCGAAACACCTTACC 1377  
Db 3991 GCGCGCGAGCG 3932  
QY 1378 GAGCGCGCGCGCGCGCATCTTCGCTGTGATGATGAGCGCGCGCGCGCGCTGTGCGAGAGAG 1437  
Db 3931 CCGCGCGCGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3872  
QY 1438 GCGTTCACGCGCGAGATGCG 1497  
Db 3871 CCGCGCGAGCG 3812  
QY 1498 CACCGCTTACATCCGCGCGTGAACGCGCGCGCGCTGCGCGCGCT 1541  
Db 3811 TACTGTCTCCCGCGCGCGCTGCGCGAGCTCACGAGCACCGCGCT 3768

RESULT 12  
US-09-105-537-1  
Sequence 1, Application US/09105537A  
Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and plikromycin  
FILE REFERENCE: 600,438US1  
CURRENT APPLICATION NUMBER: US/09/105,537A  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 15872  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-105-537-1

Query Match 4.28; Score 73.8; DB 4; Length 15872;  
Best Local Similarity 43.28; Pred. No. 7e-06;  
Matches 586; Conservative 0; Mismatches 747; Indels 24; Gaps 4;  
QY 175 CCGCGCGAGACCGCGGCTGAGAGTGTGCGCGCTGTGCGCGCGCGCTTATGACTCG 234  
Db 4799 GCGGAGAACTGCGCGCGCGCGCGAGACCGCGCTTCCCGCGCGCGCTGACAGCTGACGCG 4858  
QY 235 CACACCGAGAGAGAGCACTGCTGAGCGCGCGCGCGAGAGAGAGAGAGAGAGATGCTCCAG 294  
Db 4859 GCGCTGACAGCTGACCTGACAGCGCGCGCGCTGCGGAGATGCTGCGCGCGGAGAGACTG 4918  
QY 295 GCGCTGACAGCGGTGTGACAGCGCGAGATGCGGAGATGAGCTGCGCGCGCGCGCGCGCG 354  
Db 4919 GACCTGACCGCGGTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4978  
QY 355 AGCG 414  
Db 4979 CTGAGACAGAGCGCGCTGTGCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5038  
QY 415 CCGGAGTACCTGAGAGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 474  
Db 5039 GCGCGCGAGTGTGCG 5098  
QY 475 GCGCTTCAAGCTGCG 534  
Db 5099 GCGCGCGCTATGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5158  
QY 535 GAATCGCGCGCGAGAGCG 594  
Db 5159 GCGGAGTGTGCTGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5218  
QY 595 TCGACCGCGCGCTTCTACCG 654  
Db 5219 GAGACCGAGCGCGTGTGCG 5278  
QY 655 TCGCGCGCGGTGAGCG 714  
Db 5279 TCGCGCGAGAGCG 5338  
QY 715 CACATGCTGCGCGCGCTGAGAGAACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 774  
Db 5339 CACATGAGAGAGCTCTGACAGAGTCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5398  
QY 775 GTGATCTCGACACAGAGTATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834  
Db 5399 CCGCGGATCCCGCTGTCTACAGGTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5451  
QY 835 CCGCTGATGAGAGCG 894  
Db 5452 --GCTTACCTTCCCG 5509  
QY 895 GCGGCTCCACATGCTCAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954  
Db 5510 GCGGTCCGACCTTGGCG 5569  
QY 955 TGTGTCAAGCGCTTCCCGAGACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014  
Db 5570 GTCTTCACGAGCACTGCGCGAGAGGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 5629  
QY 1015 GCGCAATCCAAATAGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074  
Db 5630 GAGTTCACGAGTGTGCG 5689  
QY 1075 ATGAGAGAAACCGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1134  
Db 5690 GGTCTCGGAGACCG 5749  
QY 1135 GCGCTGCGCGAGAGCG 1194  
Db 5750 GCGCGCGCGAGAGCG 5809  
QY 1195 GCGGAC-----TATGCGCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1248

Db 5810 GAGGCCCTACGAGAGCCCGGCGCTGACCCGCGGGGACCGCGCTGCTGAGG 5869  
QY 1249 ATGACCGGCTGACCGCGCGCGCTTGGC-----CTGCGCGGGCGCGGACGCTGCA 1302  
Db 5870 ACCAGGTCGAGGTGCGCGCGCGGCGGCGCTGCTGACCGCGCGCTGCTGACG 5929  
QY 1303 GCCGGGTACT--TCGCGACCTGCTGCTGCTGACCGCGCGGACCGTGGCCATACCGGC 1359  
Db 5930 GACGAGCCCTGCTGCTGCGGACCATGTCACAGCGCGCGCTGCTGCGCGCGCGC 5989  
QY 1360 ACCTTGACACACCTTACGAGCGCGCGCGCGCATTCATTCGCTGACGTCACGCGCG 1419  
Db 5990 TTCTGAGAGCTGCGCTGCGCGCGCGCGGCGGACGCTGCGCGCGCTGCGGAGGACTC 6049  
QY 1420 CCGGCTGCGGACGAGCGCTTACCGCGCGGACGATCCGCGCGCGCTGCTGCGCGCG 1479  
Db 6050 ACCCTGAGAGCGCGCGCTGCTGCGCGCGCGCGCGCTGCGGATCAGAGTCCGCGG 6109  
QY 1480 GCCGCTGAGCGCGCGCGCGCGCTTACATCCGCGG 1516  
Db 6110 AGCGCGACGCGGAGTCCGCGCGCGCGCGCGCTTGC 6146

## RESULT 13

US-08-461-775-9  
Sequence 9, Application US/08461775  
Patent No. 5858773  
GENERAL INFORMATION:  
APPLICANT: MAZODIER, Philippe  
APPLICANT: GUGLIEMI, Gerard  
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,775  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/050,313  
FILING DATE: 10-MAY-1993  
APPLICATION NUMBER: FR 9011186  
FILING DATE: 10-SEP-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 010830-035  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2167 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-461-775-9

Query Match 4.2%; Score 73.4; DB 2; Length 2167;  
Best Local Similarity 46.7%; Pred. No. 6.8e-06;

Matches 233; Conservative 0; Mismatches 266; Indels 0; Gaps 0;  
QY 297 CGTACACAGGTGTGTCACGGGCAATTGCGGATCAGCTGCGCGCGCTGCGCGCGCA 356  
Db 938 CGGCCCCAAGGGCGCGGCAAGTGTGTCACGAAGATTGCGGGCGCGGACCATGACCA 997  
QY 357 CCGGCCCCCGCTGAGCTGCTGAGCAGAGCGGCTTTACCGTTTGAGCGCTTGC 416  
Db 998 CGAGGCGGTACCATATCGCCGTGAGAGGTGCAAGTCCGACGACCGCTACAGAACCTCGCGC 1057  
QY 417 CGACTACCTGAGCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCTGATGCTGG 476  
Db 1058 CCAAGCTGTCAAGAGAGGTGCGGACCAAGACCAAGCATGCGGGTGAGGACCAAC 1117  
QY 477 CCATTCAAGCTGCGCGCGCGCGGTCATGCGGACTTGCAGCGCGCGCGCGCGAGGA 536  
Db 1118 CGGACAGTGTGCG 1177  
QY 537 AATCGGCGCATGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 596  
Db 1178 CTCCCG 1237  
QY 597 GACCGCGCGCTTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656  
Db 1238 GCTCGACACCGCTC 1297  
QY 657 CCGCGCGCTGAGCG 716  
Db 1298 CCGCGAGACAGCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1357  
QY 717 CATCGTGGCGTGT 776  
Db 1358 CGGTGATCAACGTCGAGAGAGTCCACACCTTGGTGTGACCTGACTTCACCGAGG 1417  
QY 777 GATCTGCGACCAAGGTC 795  
Db 1418 CATGCGCTTGCACAGGCG 1436

## RESULT 14

US-09-031-606-9  
Sequence 9, Application US/09031606  
Patent No. 6153404  
GENERAL INFORMATION:  
APPLICANT: MAZODIER, Philippe  
APPLICANT: GUGLIEMI, Gerard  
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/031,606  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/050,313  
FILING DATE: 10-MAY-1993  
APPLICATION NUMBER: FR 9011186  
FILING DATE: 10-SEP-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113

REFERENCE/DOCKET NUMBER: 010830-035  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2167 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-09-031-606-9

Query Match 4.2%; Score 73.4; DB 3; Length 2167;  
 Best Local Similarity 46.7%; Pred. No. 6.8e-06;  
 Matches 233; Conservative 0; Mismatches 266; Indels 0; Gaps 0;

297 CGTACACAGGTGTGACAGGCAATTGGCGCATCAGCTGGCGGCGGGGACAGCCAA 356  
 Db CGGCCCCAAGGGCCGACAGTGTGATGACAGAAAGTTGGCGCCCGGACCATCACC 997  
 357 CCGCGCGCCCGCTGAGACCTGCTGAGAGAAAGCGGCTTTACCGTTTCAGCGCTTCGC 416  
 Db CGAGCGGTACACATCCCGCTGAGGTGAGTGGACAGACCGGTACAGAACTCGGCGC 1057  
 417 CGACTACCTGAGACGCTTGGCGGAGCGCGCGCGCGCTCAACCGCGCTGTATGTGG 476  
 Db CCACTCGTCAAGAGAGTGGCGACCAAGACCAACAGACATCGCGGTACCGCACACAC 1117  
 477 CCAATCAACGCTGGCGCGCGCGGTGATGCGGACTTGCAGCGCGCGCGCACAGAGGA 536  
 Db CGGACCGGTGTGCGCCAGCGCGGTGTCGCGAGGGCGCTCGCAACCTCGCGCGCGC 1177  
 537 AATCGCGCATGCGGGAAGCTGCGAGAGAGCATGCGCAAGCGCGCATCGGCATTTC 596  
 Db CTCCCGCGCGCGCTTGAAGAGGAGCATGACAGCGCGCGCTGCGCGCTGCCCGAGCT 1237  
 597 GACCGCGCGCTTCTTACCGCGCGCGCGCGCGCGCACACAGAGAGATCATGAGGTGG 656  
 Db GCTCGACACCGCGCGCGCGCGCATGAGCAAGATCCGACATCGCGCGCTGCGCGCTTC 1297  
 657 CCGCGCGGTGAGCGCGCATGCGGCGCATGACCGCACCATGCGCGAGAGAGAGAGGA 716  
 Db CGCGAGAGACAGAGGTGGGAGCTCATCGCGAGGAGAGAGAGAGAGAGAGAGGA 1357  
 717 CATCGTGGCGCGCTGAGAGAACTTCGCGATCGCGCGCGCGAGCTGAGCTCGCGTGT 776  
 Db CGGTGTATCATACGTGAGAGATCCAAACCTTCGCTGTGACCTGAGCTTACCGAGGG 1417  
 777 GATTCGACACCAAGGTC 795  
 Db 1418 CATGCGCTTCGACAGAGGC 1436

RESULT 15  
 US-08-461-775-11  
 Sequence 11, Application US/08461775  
 Patent No. 5858773  
 GENERAL INFORMATION:  
 APPLICANT: MAZODIER, Philippe  
 APPLICANT: GUGLIEMI, Gerard  
 TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Burns, Doane, Swecker & Mathis  
 STREET: George Mason Bldg., Washington & Prince Sts.  
 CITY: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/461,775  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/050,313  
 FILING DATE: 10-MAY-1993  
 APPLICATION NUMBER: FR 9011186  
 FILING DATE: 10-SEP-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Crane Feury, Sharon E  
 REGISTRATION NUMBER: 36,113  
 REFERENCE/DOCKET NUMBER: 010830-035  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2668 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-461-775-11

Query Match 4.2%; Score 73.4; DB 2; Length 2668;  
 Best Local Similarity 46.7%; Pred. No. 7e-06;  
 Matches 233; Conservative 0; Mismatches 266; Indels 0; Gaps 0;

297 CGTACACAGGTGTGACAGGCAATTGGCGCATCAGCTGGCGCGCGGACAGCCAA 356  
 Db CGGCCCCAAGGGCCGACAGTGTGATGACAGAAAGTTGGCGCCCGGACCATCACC 997  
 357 CCGCGCGCCCGCTGAGACCTGCTGAGAGAAAGCGGCTTTACCGTTTCAGCGCTTCGC 416  
 Db CGAGCGGTACACATCCCGCTGAGGTGAGTGGACAGACCGGTACAGAACTCGGCGC 1057  
 417 CGACTACCTGAGACGCTTGGCGGAGCGCGCGCGCTCAACCGCGCTGTATGTGG 476  
 Db CCACTCGTCAAGAGAGTGGCGACCAAGACCAACAGACATCGCGGTACCGCACACAC 1117  
 477 CCAATCAACGCTGGCGCGCGCGGTGATGCGGACTTGCAGCGCGCGCGCACAGAGGA 536  
 Db CGGACCGGTGTGCGCCAGCGCGGTGTCGCGAGGGCGCTCGCAACCTCGCGCGCGC 1177  
 537 AATCGCGCATGCGGGAAGCTGCGAGAGAGCATGCGCAAGCGCGCATCGGCATTTC 596  
 Db CTCCCGCGCGCGCTTGAAGAGGAGCATGACAGCGCGCGCTGCGCGCTGCCCGAGCT 1237  
 597 GACCGCGCGCTTCTTACCGCGCGCGCGCGCGCGCACACAGAGAGATCATGAGGTGG 656  
 Db GCTCGACACCGCGCGCGCGCGCATGAGCAAGATCCGACATCGCGCGCTGCGCGCTTC 1297  
 657 CCGCGCGGTGAGCGCGCATGCGGCGCATGACCGCACCATGCGCGAGAGAGAGAGGA 716  
 Db CGCGAGAGACAGAGGTGGGAGCTCATCGCGAGGAGAGAGAGAGAGAGAGAGGA 1357  
 717 CATCGTGGCGCGCTGAGAGAACTTCGCGATCGCGCGCGCGAGCTGAGCTCGCGTGT 776  
 Db CGGTGTATCATACGTGAGAGATCCAAACCTTCGCTGTGACCTGAGCTTACCGAGGG 1417  
 777 GATTCGACACCAAGGTC 795  
 Db 1418 CATGCGCTTCGACAGAGGC 1436

Search completed: May 11, 2003, 10:52:38  
 Job time: 235 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 10:48:56 ; Search time 230 Seconds  
(without alignments)  
9498.588 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 1758  
Sequence: 1 gaattccactgacgcgcga.....ccctagactcagaagcct 1758

Scoring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 783854 segs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	86	4.9	88421	US-09-976-059-1 Sequence 1, Appl1
2	81.8	4.7	1248	US-09-860-846-7 Sequence 7, Appl1
3	81.8	4.7	1248	US-09-988-384B-7 Sequence 7, Appl1
4	81.8	4.7	1248	US-09-861-289-7 Sequence 7, Appl1
5	81.8	4.7	12441	US-09-988-384B-3 Sequence 3, Appl1
6	81.8	4.7	13613	US-09-860-846-3 Sequence 3, Appl1
7	81.8	4.7	13613	US-09-861-289-3 Sequence 3, Appl1
8	79.2	4.5	4341	US-10-166-087-45 Sequence 45, Appl1
9	79.2	4.5	32539	US-09-796-679-5 Sequence 5, Appl1
10	75.8	4.3	4020	US-09-860-846-1 Sequence 1, Appl1
11	73.8	4.2	15872	US-09-988-384B-1 Sequence 1, Appl1
12	73.8	4.2	15872	US-09-861-289-1 Sequence 1, Appl1
13	73.8	4.2	15872	US-09-861-289-1 Sequence 1, Appl1
14	73.4	4.2	4257	US-09-825-288A-1 Sequence 1, Appl1
15	72.4	4.1	3189	US-09-815-242-4056 Sequence 4056, Ap
16	72.2	4.1	13842	US-09-860-846-30 Sequence 30, Appl
17	72.2	4.1	13842	US-09-988-384B-30 Sequence 30, Appl
18	72.2	4.1	13842	US-09-861-289-30 Sequence 30, Appl
19	72.2	4.1	36778	US-09-860-846-5 Sequence 5, Appl1

20	72.2	4.1	36778	US-09-861-289-5 Sequence 5, Appl1
21	72.2	4.1	37948	US-09-988-384B-5 Sequence 5, Appl1
22	69.6	4.0	11220	US-09-860-846-32 Sequence 32, Appl
23	69.6	4.0	11220	US-09-988-384B-32 Sequence 32, Appl
24	69.6	4.0	11220	US-09-861-289-32 Sequence 32, Appl
25	67.8	3.9	1014	US-09-860-846-13 Sequence 13, Appl
26	67.8	3.9	1014	US-09-988-384B-13 Sequence 13, Appl
27	67.8	3.9	1014	US-09-861-289-13 Sequence 13, Appl
28	67.8	3.9	13029	US-09-815-242-4052 Sequence 4052, Ap
29	64.2	3.7	2712	US-09-748-033-4 Sequence 4, Appl1
30	63.2	3.6	1764	US-09-815-242-7884 Sequence 7884, Ap
31	62.4	3.5	50937	US-09-808-880-1 Sequence 1, Appl1
32	62.2	3.5	804	US-09-773-748-2 Sequence 2, Appl1
33	62.2	3.5	1290	US-09-815-242-7834 Sequence 7834, Ap
34	61.6	3.5	88421	US-09-976-059-1 Sequence 1, Appl1
35	60.8	3.5	1185	US-09-887-576-784 Sequence 784, App
36	60.8	3.5	2787	US-09-860-846-40 Sequence 40, Appl
37	60.8	3.5	2787	US-09-988-384B-40 Sequence 40, Appl
38	60.8	3.5	2787	US-09-861-289-40 Sequence 40, Appl
39	60.4	3.4	4689	US-09-860-846-34 Sequence 34, Appl
40	60.4	3.4	4689	US-09-988-384B-34 Sequence 34, Appl
41	60.4	3.4	4689	US-09-861-289-34 Sequence 34, Appl
42	60.2	3.4	1128	US-10-318-142-1 Sequence 1, Appl1
43	60	3.4	7419	US-09-815-242-4009 Sequence 4009, Ap
44	59.8	3.4	1815	US-09-758-269-13 Sequence 13, Appl
45	59.8	3.4	1839	US-09-815-242-4006 Sequence 4006, Ap

## ALIGNMENTS

RESULT 1  
US-09-976-059-1  
Sequence 1, Application US/09976059  
Patent No. US20020164747A1  
GENERAL INFORMATION:  
APPLICANT: Farnet, Chris  
APPLICANT: Zazopoulos, Emmanuel  
APPLICANT: Staiffa, Alfredo  
TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
FILE REFERENCE: 3019-PCP  
CURRENT APPLICATION NUMBER: US/09/976, 059  
CURRENT FILING DATE: 2001-10-15  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 88421  
TYPE: DNA  
ORGANISM: Actinoplanes sp.  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (2077)..(3078)  
OTHER INFORMATION: ORF 1; positive strandedness  
NAME/KEY: misc.feature  
LOCATION: (3118)..(4032)  
OTHER INFORMATION: ORF 2; positive strandedness  
NAME/KEY: misc.feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3; positive strandedness  
NAME/KEY: misc.feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3; positive strandedness  
NAME/KEY: misc.feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3; positive strandedness  
NAME/KEY: misc.feature  
LOCATION: (6665)..(5814)  
OTHER INFORMATION: ORF 4; negative strandedness  
NAME/KEY: misc.feature  
LOCATION: (7703)..(6693)  
OTHER INFORMATION: ORF 5; negative strandedness  
NAME/KEY: misc.feature  
LOCATION: (9464)..(8130)  
OTHER INFORMATION: ORF 6; negative strandedness  
NAME/KEY: misc.feature  
LOCATION: (9691)..(10761)





Db 47124 TCCAGGTCATGTCACCTCGAGAGGCGGCGGCGGAGCGGCGGAGCTGCGGGGCTGG 47183  
QY 989 ACCTGATGATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1048  
Db 47184 CGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 47243  
QY 1049 AGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1108  
Db 47244 CGAGACCTTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47303  
QY 1109 TCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1168  
Db 47304 CGACCTTCGTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 47363  
QY 1169 TGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1228  
Db 47364 GCGTCGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 47411  
QY 1229 TGGAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1288  
Db 47412 TGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47471  
QY 1289 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1348  
Db 47472 CGGAGATCACCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 47531  
QY 1349 CGGATACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1408  
Db 47532 CGGTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47591  
QY 1409 TCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47615  
Db 47592 TCGCCGCGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47615

RESULT 2  
US-09-860-846-7  
; Sequence 7, Application US/09860846  
; Patent No. US20020164742A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D. H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/860.846  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PastSeq for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 1248  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-860-846-7

Query Match 4.7%; Score 81.8; DB 9; Length 1248;  
Best Local Similarity 44.2%; Pred. No. 1.4e-09;  
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

QY 32 CCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 91  
Db 215 CCACCTGCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 274  
QY 92 AGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 151  
Db 275 AAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 334  
QY 152 CCAATGCGGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 211

Db 335 TCACCCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 394  
QY 212 TCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 271  
Db 395 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 454  
QY 272 ACATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 331  
Db 455 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 514  
QY 332 GCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 391  
Db 515 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 574  
QY 392 GCTTTCACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 451  
Db 575 AGGTCCTTCAAGTTTCA---GCGACGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631  
QY 452 CCGTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 511  
Db 632 TCACCGACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 691  
QY 512 TGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571  
Db 692 TGCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 751  
QY 572 TGGCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631  
Db 752 TGGGCGTCACTCCCTTCGACGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 811  
QY 632 CCAACGGAAGATCATGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 691  
Db 812 CCGCTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 871  
QY 692 CCACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 751  
Db 872 ACGGCTCAAC 931  
QY 752 GCGCGAGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 811  
Db 932 ACCGCGACCTCGTCATGAGAGTCTGTAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 991  
QY 812 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 862  
Db 992 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1051  
QY 863 AGGACGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 922  
Db 1052 CCGAACGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1111  
QY 923 GCGTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 982  
Db 1112 ACATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1171  
QY 983 GCGCGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1015  
Db 1172 CCGGCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1204

RESULT 3  
US-09-988-384B-7  
; Sequence 7, Application US/09988384B  
; Publication No. US20030073824A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D. H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.536US1  
; CURRENT APPLICATION NUMBER: US/09/988.384B  
; PRIOR FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: PCT/US99/14398



Query Match	4.7%	Score 81.8	DB 9	Length 12441
Best Local Similarity	44.2%	Pred. No. 1.1e-09		
Matches 439	Conservative 0	Mismatches 542	Indels 12	Gaps
QY 32	CCATGTCCCATCCGATTC	CCAGCCCTTCGACGCTGTC	TCGCGGGGGGGAGACCCCTCATG	91
DB 11982	CCACCTGCACACGCCAC	GGCGGGCTCCACGCTCTCGCGC	ACACCGCCCGGGCTCACCGGGG	11933
QY 92	AAGGAGACAACACCCCG	GGGGGGCGCGCCGACCTG	GGGGCGGACCGCATGCGCC	151
DB 11922	AAGTATATATGCGCTG	ATGATGACGTTTCGCGCGAC	CCCGCACCGACATTCGCGTGGATCGGCG	11863
QY 152	CCATGCGGATCTGTCG	AGACCGCGCGACACCGGGG	TCGACGCTGTGGGGCTGTGG	211
DB 11862	TCACCCGGTCTTGCCG	ACATTCACACCCCGGAC	ACCGGCAACTCTGACCGGGACAGATGG	11803

RESULT 6  
 US-09-860-846-3/C  
 Sequence 3, Application US/09860846  
 Patent NO. US20020164742A1  
 GENERAL INFORMATION:  
 APPLICANT: Sherman, D.H.  
 APPLICANT: Liu, H.  
 APPLICANT: Xue, Y.  
 APPLICANT: Zhao, L.  
 TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 FILE REFERENCE: 600,438051  
 CURRENT APPLICATION NUMBER: US/09/860,846  
 CURRENT FILING DATE: 2001-05-18  
 PRIOR APPLICATION NUMBER: 09/105,537  
 PRIOR FILING DATE: 1998-06-26  
 NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 3  
 LENGTH: 13613  
 TYPE: DNA  
 ORGANISM: Streptomyces venezuelae  
 US-09-860-846-3

Query Match 4.7%; Score 81.8; DB 9; Length 13613;  
 Best Local Similarity 44.2%; Pred. No. 1.1e-09;  
 Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

```

OY 32 CCATGTCCTCCATCCGATTCCTCCAGCCCTTCGACCTCTCTCTCGCGGCGGCAACCTTCATCG 91
DB 12669 CCACCTGCAAGCGCACGCGCGCGGCTCCAGCTCTCTCTCGCGGCGGCGGCTTCACCGGCG 12610
OY 92 ACGGAGCAACACCCCGGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 151
DB 12609 AAGTATCATGCGCGCGATGAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12550
OY 152 CCATCGGCGATCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 211
DB 12549 TCACCGCGGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12490
OY 212 TCGGCGCGGCGGCTTCATGACCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 271
DB 12489 CCGCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12430
OY 272 ACATGACGCCCAAGATCTCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 331
DB 12429 GCGCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12370
OY 332 GCTTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 391
DB 12369 CCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12310
OY 392 GCTTTCACGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 451
DB 12309 AGGTCTTCACGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12253
OY 452 CCGTCAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 511
DB 12252 TCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12193
OY 512 TGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571
DB 12192 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12133
OY 572 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 631
DB 12132 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12073
OY 632 CCACGAGAGATCATGAGGCTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 691
DB 12072 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12013
OY 692 CCACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 751
DB 12012 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 11953
OY 752 GCGCGGAGGCTGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 811
DB 11952 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 11893
OY 812 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 862
DB 11892 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11833
OY 863 AGGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 922
DB 11832 CCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11773
OY 923 GCGTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 982

```

```

DB 11772 ACATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11713
OY 983 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1015
DB 11712 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11660.

```

RESULT 7  
 US-09-861-289-3/c  
 Sequence 3, Application US/09861289  
 Patent No. US20020110897A1  
 GENERAL INFORMATION:  
 APPLICANT: Sherman, D.H.  
 APPLICANT: Liu, Y.  
 APPLICANT: Xue, Y.  
 APPLICANT: Zhao, L.  
 TITLE OF INVENTION: DNA encoding methymycin and plixromycin  
 FILE REFERENCE: 600,438US1  
 CURRENT APPLICATION NUMBER: US/09/861,289  
 PRIOR FILING DATE: 2001-05-18  
 PRIOR APPLICATION NUMBER: 09/105,537  
 NUMBER OF SEQ ID NOS: 43  
 NUMBER OF SEQ ID NOS: 43  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 3  
 LENGTH: 13613  
 TYPE: DNA  
 ORGANISM: Streptomyces venezuelae  
 US-09-861-289-3

Query Match 4.7%; Score 81.8; DB 10; Length 13613;  
 Best Local Similarity 44.2%; Pred. No. 1.1e-09;  
 Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

```

OY 32 CCATGTCCTCCATCCGATTCCTCCAGCCCTTCGACCTCTCTCTCGCGGCGGCGGCGGCGGCGGCG 91
DB 12669 CCACCTGCAAGCGCACGCGCGCGGCTCCAGCTCTCTCTCGCGGCGGCGGCGGCGGCGGCGGCG 12610
OY 92 ACGGAGCAACACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 151
DB 12609 AAGTATCATGCGCGCGATGAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12550
OY 152 CCATCGGCGATCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 211
DB 12549 TCACCGCGGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12490
OY 212 TCGGCGCGGCGGCTTCATGACCTCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 271
DB 12489 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12430
OY 272 ACATGACGCCCAAGATCTCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 331
DB 12429 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12370
OY 332 GCTTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 391
DB 12369 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12310
OY 392 GCTTTCACGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 451
DB 12309 AGGTCTTCACGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12253
OY 452 CCGTCAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 511
DB 12252 TCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12193
OY 512 TGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571
DB 12192 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12133
OY 572 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 631
DB 12132 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12073

```



```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 32539
; TYPE: DNA
; ORGANISM: Streptomyces refuineus subspecies thermotolerans
US-10-166-087-1

```

## Query Match

```

Best Local Similarity 41.8%; Score 79.2; DB 9; Length 32539;
Matches 474; Conservative 0; Mismatches 658; Indels 0; Gaps 0;

```

```

QY 31 TCCATGTCCTCCATCCGATTCCTCAAGCCCTTCGACCTGCTCTGCGGGCGGACCCCTCATC 90
DB 25011 TCGCTGTCCCTGCTGCTGACAGAGTGGCGGACGGCCCGCGGGCGGCGGCTGCTGCTG 25070
QY 91 GAGGCGAGCAGACACCCCGGGGGGGGGCGGCGGACCTGGGGCTGGGGCGGACCGCATCGCC 150
DB 25071 CCGGGGGAGAGGGCCCGCGGGCGGACGCTGACACCAAGAGGGCTCTGCGGAGCGGCTG 25130
QY 151 GGCATCGCGGATCTGTGAGACCGCGCGGACACCCCGGGTGCAGCTGTGCGGGCTGGTG 210
DB 25131 GCGGAGTGAAGACCTTGTGCGGCACGAGGGCATCACACCGGATGTTCTTCCGCGCC 25190
QY 211 GTGCGCGCGCGGCTTCATGAGCTCGGACCCACGACGACAACTACTGCTCAGGCGTGC 270
DB 25191 CTGCGCTTCGCTGCTGCTGACACCGGACACCGGAGGACGACATCTCTCTGCGGGTCCG 25250
QY 271 GACATGACGCGCCCAAGATCTGCGAGGGGCTCACACGAGTGTGCAGGGGCAATTGCGGCATC 330
DB 25251 GTCAACGAGGGGGGAGGGGAGCGGACCGGAGGTCTCGGGCACCTGACCAACACGGTCTG 25310
QY 331 AGCTGCGCGCGCTGCGGCGACCGCACCCCGCGCGCGCGCTGAGACCTGCTGAGAGAGGC 390
DB 25311 CTGCGGACACGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25370
QY 391 GGCCTTACCGCTTCGAGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 450
DB 25371 GACATGCTCAACGCGCTGCGGCGACCGGACATGCTCCGCTGAGGGGAGTGTGCGGACATC 25430
QY 451 GCGGTCAACGCGCGCTTCGAGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTC 510
DB 25431 CCGGCGCGCTGGAGAGCGGCGGACGAGCGGCTGCGGACCTGTTCAACGAGTGTCTACG 25490
QY 511 TTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570
DB 25491 ATGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25550
QY 571 ATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630
DB 25551 GTCTCGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25610
QY 631 ACCACGAGAGATCATCGAGGTGTGCGCGCGCTGAGGGCGGCGCATCTTACGCC 690
DB 25611 ACGCTGTGCTGAGAGACACCTCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25670
QY 691 ACCACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750
DB 25671 GCGCGCGCGCTGAGACGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25730
QY 751 GCGCGCGAGTGTGAGCTGCGCGGTGTGATCTGACACCAAGTATGAGCGCGCGCAT 810
DB 25731 CCGCTGCGCTGTGAGCGCGGAGGAGGAGCGGCTGTACCGGCTGTGCGCGCGCGAGG 25790
QY 811 TTGCGCGCGCTGCGCGAGACGCTGCGCGGTGTGATCGAGGGCGCGCATGGCGCGCGAGAG 870
DB 25791 CAGGACGCGCGCGAGCTGGGACGAGGTGACGCGCGCGCATGTTGCGCGAGCGCGCGCG 25850
QY 871 TCGCTGAGCGCGTATCCCTTACGTGCGCGGCTCCACATGCTCAAGAGAGCGGCTGTG 930
DB 25851 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25910
QY 931 CTGCGCGGAGCGACATCATATCACTGTGTGCAAGCGCTTCCCGGAACTGAGCGCGCGCGAG 990

```

```

DB 25911 GCGCGGACAGCCGACCGCGTGGCGGGGAGCTGCGCGCGCGGAGTGGCGGACGCGG 25970
QY 991 CTGATGAGTGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1050
DB 25971 CCGGTGCGCGGTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26030
QY 1051 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1110
DB 26031 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26090
QY 1111 GCGCGGACCATGATGCGCTCCGACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1162
DB 26091 GCGGTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26142

RESULT 10
US-09-796-679-5/C
; Sequence 5, Application US/09796679
; Publication No. US20030013076A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Anthony J
; APPLICANT: Lytle, David J
; TITLE OF INVENTION: University of Otago
; TITLE OF INVENTION: Parapoxvirus vectors
; FILE REFERENCE: 23607 MRB
; CURRENT APPLICATION NUMBER: US/09/796,679
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/155,421
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: PCT/NZ97/00040
; PRIOR FILING DATE: 1997-03-27
; PRIOR APPLICATION NUMBER: NZ 286284
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4020
; TYPE: DNA
; ORGANISM: Orf virus strain NZ-2
US-09-796-679-5

Query Match
Best Local Similarity 44.3%; Score 75.8; DB 9; Length 4020;
Matches 434; Conservative 0; Mismatches 527; Indels 12; Gaps 3;

```



QY 780 CTGCACACACAGGTCATGGGCGACGCCAATTTCGGCCGCTCGGCGGAGACCTGCCCT 839  
Db 3289 CGGGAACCCAAAGCCCTCCGACATGCAACGGCTGACCGCGCTGCGGCTGTGTCT 3230  
QY 840 GATCGAGGCGCCATGAGCGCCGAGACGTCCTGCTGAGACGCTATCCCTACGTGGCCG 899  
Db 3229 CGGCGGCGTCTCGCGCGCGCTGCTGAGACATGATCTCGCGCGCGCGCGCGCTGCT 3170  
QY 900 CTCACCATCTCTAAGACGAGACCGCGCTGCTGCTGCGGAGACACATCATCCTGGTG 959  
Db 3169 CGTGGAGCGCTAGCGGCGCAACAGCTGACACACCTCGCGCGCGCGCAAGATCTCGA 3110  
QY 960 CAAGCCCTTCCCGACATGAGCGGCGGCGACCTGAGTAAAGTCCGCGCGCGCGCA 1019  
Db 3109 GGGCATGCTGCGACGCTCAGCGGCTTACGCGTGA-----CCGCGCGCGCTGAGCGC 3056  
QY 1020 ATCCAAATACAGACGTGGTCCGAGCTGACCGCGCGCGCGCTTCTATGATGGA 1079  
Db 3055 GTGTGGGAACACCATGTGCTGACTACATGGGAGCTACGGAGCGCGCGCTGCT 2996  
QY 1080 CGAACCCGACGTGACGCGCTCTGCGCTGCGCGCGCGACATGATGCTCGACGCGCT 1139  
Db 2995 GGACTTGTGCTGAGCGCGCGCTGACCTGAACTGCGCAACAACTGACAGCGC 2936  
QY 1140 GCGCGACGACGACCGCGCATCCGCGCTGAGGCGACCTTCCGCGCGCTGCGGCGCA 1199  
Db 2935 GCGGACCGCGCGC---CGGCTGTCAGCCACGCGCGCTGCTGCGCGCTGCTGAGG 2879  
QY 1200 CTATGCGCGGACCTGCGCTGCTTCCGCTGAGACCGCGCTGATGAAATGACCGCGCT 1259  
Db 2878 CGGCGAGCTGAGACACATGTCGCGCTGCGCGCTATGCGCGGCTTCCGAGATGCTGCGCG 2819  
QY 1260 GACCGCGCGCGCTGCGCTGCGCTGCGCGCGCGCGCGCGCGCTGATGAAATGACCGCG 1319  
Db 2818 CAAACAAGTGGCG 2759  
QY 1320 CCTGTGTGTGTTT 1332  
Db 2758 GCGCGCGCTGATC 2746

RESULT 11  
US-09-860-846-1  
; Sequence 1, Application US/09860846  
; Patent No. US2002016742A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D. H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and plikromycin  
; FILE REFERENCE: 600,438US1  
; CURRENT APPLICATION NUMBER: US/09/860,846  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version. 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-860-846-1

Query Match 4.2%; Score 73.8; Db 9; Length 15872;  
Best Local Similarity 43.2%; Pred. No. 6.3e-08;  
Matches 586; Conservative 0; Mismatches 747; Indels 24; Gaps 4;

QY 175 GCGCGGACACCGGCTGACGCTGCGGCTGCTGCTGCGCGCGCGCTTATCGACGCG 234  
Db 4799 GCGCGGACACCGGCTGACGCTGCGGCTGCTGCTGCGCGCGCGCTTATCGACGCG 4858  
QY 235 CACACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 294

Db 4859 GCCCTCGACGCTACCTGACCGCGCGCGCGGAGATGCTGCGCGCGGAGACGCTG 4918  
QY 295 GCGCTACACGAGGTGCTGACGCGGCAATTGCGGATACGCTGCGCGCGCGCGCGCGCG 354  
Db 4919 GACCTACCGGTAACACGACCGCGCGCGCTTCTGCGCTTCTGAGGCGCGCTGCGCGCT 4978  
QY 355 AACCG 414  
Db 4979 CTGCAACACACG 5038  
QY 415 GCGGACGACGCTGAGCGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 474  
Db 5039 GCG 5098  
QY 475 GCGCATTCACG 534  
Db 5099 GCG 5158  
QY 535 GAATCCG 594  
Db 5159 GCGGAGGTCGCTGAGTCCCTGAGAGGCTACGAGGCGCGCGCGCGCGCGCGCGCGCG 5218  
QY 595 TCGACCGCGCGCTTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 654  
Db 5219 GAGCCACCG 5278  
QY 655 TCG 714  
Db 5279 TGGCGGAGACCG 5338  
QY 715 CACATCTGCG 774  
Db 5339 CACATGACGACGCTGCTGAGAGGTTCTCTCGGCTCCGAGGCGCGCGCGCGCGCGCG 5398  
QY 775 GTGATCTGCAACCAAGGTCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834  
Db 5399 CCGCGGATCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5451  
QY 835 CCGGTATGAGGCG 894  
Db 5452 --GCTACCTGCG 5509  
QY 895 GCGCGCTCACATGCTGCAAGACGAGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 954  
Db 5510 GCGGTCCGACCGCTGCG 5569  
QY 955 TGTGCAAGCGCTTCCCGCAATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014  
Db 5570 GTCTCAGCGCACTCGCGCGAGAGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5629  
QY 1015 GCGCAATCCAAATGACAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074  
Db 5630 GATGTCAGCGGTCTCCCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5689  
QY 1075 ATGACGACCG 1134  
Db 5690 GGTCTGCGACCG 5749  
QY 1135 GCGCTGCG 1194  
Db 5750 GCG 5809  
QY 1195 GGGGAC-----TATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1248  
Db 5810 GAGGCG 5869  
QY 1249 ATGACCG 1302  
Db 5870 ACCAGCGTGAAGTGGCG 5929  
QY 1303 GCGGCGTACT---TCGCGACCTGTGTGATGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1359

Db 5930 GACGACCCCTGCTGGCGGACACATGCTCAACGGCGCCGCTCTGTTGCGGCGACCGCC 5989  
 QY 1360 ACCTTCGAACACCCCTACGAGCGCGCGCCGCGCATTCCTGCTGATGCTAAGCGCGC 1419  
 Db 5990 TTCTTGAGAGCTGCGCTTGGCGCGGCGACCATCTGGGGGGGCTCCGGGTGAGAGAACTC 6049  
 QY 1420 CCGGCTTGGAAGCAAGGCGCTTACCGCGGACAGCATGGCGCGCGGCTGCTGCAAGCGACG 1479  
 Db 6050 ACCCTCGAAGCGCGGCTGCTGCTGCTCCGAGAGCGGGGCGCTCCGCAATCAGTGGCGGTG 6109  
 QY 1480 GCGGCTGAGCGCGCGCGCGCGACCTTACATCCGCGC 1516  
 Db 6110 AGCGGCGAGCGGAGTGGCGCGCGCGCGCGACCTTTCG 6146

## RESULT 12

US-09-988-384B-1  
 ; Sequence 1, Application US/09988384B  
 ; Publication No. US20030073824A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600,536US1  
 ; CURRENT APPLICATION NUMBER: US/09/988,384B  
 ; PRIOR FILING DATE: 1999-06-25  
 ; PRIOR APPLICATION NUMBER: PCT/US99/14398  
 ; PRIOR FILING DATE: 1999-06-25  
 ; PRIOR APPLICATION NUMBER: US 09/105,537  
 ; PRIOR FILING DATE: 1998-06-26  
 ; NUMBER OF SEQ ID NOS: 53  
 ; SEQ ID NO 1  
 ; LENGTH: 15872  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-988-384B-1

## Query Match

Best Local Similarity 4.2%; Score 73.8; DB 9; Length 15872;  
 Matches 586; Conservative 0; Mismatches 747; Indels 24; Gaps 4;

QY 175 GCGCGCACACCCGGTGCAGTGTGCGGCTGTGTCGCGCGCGCGCTTCATGACTCG 234  
 Db 4799 GCGGAGGAATGTCGGCGCGCGCGACCCCGTTCGCGCGCGCGCTTCGACACCGTGTGCGC 4858  
 QY 235 CACACCCAGACGACAACTACTGCTCAGGCGTGGGACATGACGCCCAAGATCTCGCAG 294  
 Db 4859 GCGCTGACCGTCACTCGACCGCGCGCGCGGAGATGCTGCGCGCGCGCGGAGAGCTG 4918  
 QY 295 GCGGTACCAAGTGTGTCAGCGGCAATTCGCGGATAGCGTGGCGCGCGCTGCGGACGCC 354  
 Db 4919 GACCTACCGCGGACACCGACCGCGCGCTTCCTGCGCTTCGAGAGTGGCGGTTCGCGCTC 4978  
 QY 355 AACCCCGCGCGCGCGCTGACCTGTGGAAGAGCGGCGCTTACCGGCTTCGAGCGCTTC 414  
 Db 4979 CTGGAACACACAGCGCTTCCTCCCGACGCTGTACCGCGCACCTCCCTCGCGGAGTGGCG 5038  
 QY 415 GCGGACTACCTGGAGCGGTTGGGGCGCACCGCGCGCGCTCAACCGCGCTGTATGCTG 474  
 Db 5039 GCGGCGCACGTCGGCTGCTCTCTCTCGACGACGCGCGGACGCTGCTCAACCGCGCGC 5098  
 QY 475 GCGCATTAACGCTGGCGCGCGCGCTGATGCGGACTTCAGAGCGCGCGCGCGCGACGAG 534  
 Db 5099 GCGCGGCTATGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 5158  
 QY 535 GAATTCG 594  
 Db 5159 GCGGAGTGTGAGTCCCTGGAAGGCTTCAAGAGGCGAGGAGTGGCGGTTCGCGCGCGCGTCAAC 5218  
 QY 595 TGACCGGCGCGCTTACCG 654

Db 5219 GACCCACACCGCGCTGCTGCTTCGCGCGAGACCGCGCGCGCGCGGAGAGATCCGCGCTA 5278  
 QY 655 TGCGCGCGCTGAGCGCGCATATGCGGCACTTACGCCACCCACATGCGGAGAGAGCGAG 714  
 Db 5279 TGCGCGGAGACCGCGCGCGCGCGCACCGCGAGGCTGCGCGCTGACGACGCGCTTCACCTCCG 5338  
 QY 715 CACATCGTGGCGCGCTGAGAGAACTTCGCAATCGCGCGCGCGAGCTGAGCTGCGGTG 774  
 Db 5339 CACATGAGCAGCTCTGAGAGATTCCTCCGGGTGCGCGAGGCGCTACCTTGGAGAG 5398  
 QY 775 GTGATTCGACACCAAGATATGCGGCGGACCCCAATTTGCGCGCTGCGCGAGAGCTG 834  
 Db 5399 CCGGAGATCCCGCTGCTCTCCAGCTTACCGCGCGCGCTGCTGAGTCCGCGGCA----- 5451  
 QY 835 CCGCTGATGAGCGCGCGCATGCGCGCGAGAGAGCTGCTGCGGAGCGGATCCCTAGCTG 894  
 Db 5452 --GCTCACTGCGCGCGCGGCTGAGTGGGTGACCAATCCGCGCGCGCGCTTCCTGAGAC 5509  
 QY 895 GCGGCTTCACCATGCTCAAGCAGAGACCGCGCTGCTGCGCGGAGCGCACCATCATCAGC 954  
 Db 5510 GCGCTCCGACACCTGCGCGCGCGCGCGAGAGCGAGCGCTCTGCTGAGATCGCGCGCGAGCC 5569  
 QY 955 TGGTCAAGCGCTTCCCGGAACTGAGCGGCGCGCGACCTGATGAATCGCGCGCGAGCGC 1014  
 Db 5570 GTCTTCAGGCACTGCGCGGAGAGAGCTGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCG 5629  
 QY 1015 GCGCAATTCAGTACGAGCTGCTGCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCATCTTATG 1074  
 Db 5630 GAGGTACAGGTGCTGCG 5689  
 QY 1075 ATGAGCAGACCCGAGCTGACAGCGCATCTGCGCTGCGCGCGCGCGCGCATGATGCTCGAC 1134  
 Db 5690 GTCTTCGCGACCGAC 5749  
 QY 1135 GCGCTGCGCGACGAGCG 1194  
 Db 5750 GCGCGCGCGACGAGCG 5809  
 QY 1195 GCGCGAC-----TATGCGCGCGAGCTGCGCGCGCTTCCCGCTGAGAGCGGCGATGAGAG 1248  
 Db 5810 GAGCGCGCGTACGAGAGCG 5869  
 QY 1249 ATACCGGCTTACCG 1302  
 Db 5870 ACCACGCTGAGGTGCG 5929  
 QY 1303 GCGGCGACT---TGGCGGACGCTGCTGCTGCTTGTGAGACCGCGCGCGCGCGCGCGCGCG 1359  
 Db 5930 GACCAACCGCTGCTGCG 5989  
 QY 1360 ACCTTCGAACACCCCTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1419  
 Db 5990 TTCTTGAGACTGCGCGCTTCG 6049  
 QY 1420 CCGGCTTGCAAGAGAGCGGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1479  
 Db 6050 ACCCTCGAAGCGCGCGCTGCTCTGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 6109  
 QY 1480 GCGGCGTGAAGCG 1516  
 Db 6110 AGCGGCGAGCGGAGTGGCG 6146

## RESULT 13

US-09-861-289-1  
 ; Sequence 1, Application US/09861289  
 ; Patent No. US2002011087A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600.438US1  
 CURRENT APPLICATION NUMBER: US/09/861,289  
 CURRENT FILING DATE: 2001-05-18  
 PRIOR APPLICATION NUMBER: 09/105,537  
 PRIOR FILING DATE: 1998-06-26  
 NUMBER OF SEQ ID NOS: 43  
 SOFTWARE: FASTSEQ for Windows Version 3.0  
 SEQ ID NO 1  
 LENGTH: 15872  
 TYPE: DNA  
 ORGANISM: Streptomyces venezuelae  
 US-09-861-289-1

Query Match 4.2%; Score 73.8; DB 10; Length 15872;  
 Best Local Similarity 43.2%; Pred. No. 6.3e-08;  
 Matches 386; Conservative 0; Mismatches 747; Indels 24; Gaps 4;

175 GCGCGACACACCGGCTGACGCTGTCGCGGCTGTCGCGGCTTCATGACGCTG 234  
 4799 GCGGAGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4858  
 235 CACACCCACGACGACAACTGCTGTCGAGCGCTGCGACATGACGCCAAGATCTG 294  
 4859 GCGCTGACCGCTGACCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4918  
 295 GCGGTACACAGGCTGCTGACGCGGCAATGCGGCTGATGACCTGCGCGCGCGCG 354  
 4919 GACCTGACCGCTGACACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4978  
 355 AACCG 414  
 4979 CTGGAACACACG 5038  
 415 GCGGACCTGTCGACGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 474  
 5039 GCGCGGACGCTGCG 5098  
 475 GCGCATTCACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534  
 5099 GCGCGGCTGATGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5158  
 535 GAAATCG 594  
 5159 GCGGAGGTGCTGAGTCCCTGGAAGGCTGACGAGGCGAGGCTGCGCGCGCGCG 5218  
 595 TCGACCGCGCGCTTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 654  
 5219 GACCCGACCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5278  
 655 TCGCGCGCGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 714  
 5279 TCG 714  
 715 CACATGTCG 774  
 5339 CACATGAGAGAGTCTCTGACAGATCTCTCGGCTGCGCGCGCGCGCGCGCGCG 5398  
 775 GTGATCTGCGCACACAGGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834  
 5339 CCGCGGATGCGCGCTGCTGCAAGGCTGACCGCGCGCGCGCGCGCGCGCGCGCG 5451  
 835 CCGGCTGATGAGAGCG 894  
 5452 -GCTGACCTGCG 5509  
 895 GCGGCTGACACGCTGCTGACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954  
 5510 GCGGCTGCGACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5569  
 955 TGGTCAAGACCGCTTCCCGCACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014  
 5570 GTCTGACGCGACCTGCGCGAGAGGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCG 5629

1015 GCGAATTCAGATGACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074  
 5630 GACCTGACGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5689  
 1075 ATGAGCAACCG 1134  
 5690 GTCTGCG 5749  
 1135 GCGCTGCG 1194  
 5750 GCGGCG 5809  
 1195 GCGGAC-----TATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1248  
 5810 GAGGCG 5869  
 1249 ATGACG 1302  
 5870 ACAGGCTGAGGCTGCG 5929  
 1303 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1359  
 5930 GACAGCG 5989  
 1360 ACCTGCAACACCTTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1419  
 5990 TTCTGAGGCTGCG 6049  
 1420 CCGGTGCGCGAGAGCG 1479  
 6050 ACCCTGAGCG 6109  
 1480 GCGGCTGAGCG 1516  
 6110 AGCGGCGAGCGCGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6146

RESULT 14  
 US-09-825-288A-1  
 ; Sequence 1, Application US/09825288A  
 ; Publication No. US20020192822A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEOPARDI, ROSARIO  
 ; APPLICANT: ROIZMAN, BERNARD  
 ; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS  
 ; FILE REFERENCE: ARCD:317US01  
 ; CURRENT APPLICATION NUMBER: US/09/825,288A  
 ; CURRENT FILING DATE: 2001-04-02  
 ; PRIOR APPLICATION NUMBER: 09/259,821  
 ; PRIOR FILING DATE: 1999-03-01  
 ; PRIOR APPLICATION NUMBER: 08/690,473  
 ; PRIOR FILING DATE: 1996-07-26  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 4257  
 ; TYPE: DNA  
 ; ORGANISM: HERPES VIRUS, TYPE 1  
 ; US-09-825-288A-1

Query Match 4.2%; Score 73.4; DB 9; Length 4257;  
 Best Local Similarity 42.6%; Pred. No. 8.9e-08;  
 Matches 581; Conservative 0; Mismatches 766; Indels 18; Gaps 3;

189 GGTGACGCTGTCG 248  
 1548 GGTGCGCGGAGAGTGGCGCTGACCGAGCGCTGCTGCGGAGTCTGCGCGCGCG 1607  
 249 CAACTGCTGTCAGCGCTGCGGACATGACGCCCAAGATCTGCGAGGCGGTACACG 308  
 1608 CAGACGCTCTCTATACACGCGGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCG 1667  
 309 GGTGACGCGCAATGTCGCGATCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368

```

Db 1668 GGGGCGGCGCGCTTCCGGCTGGGGCGCTGGCGACGCGCGCGCGCGCTTGGCCATGAGCCG 1727
QY 369 CTGAGACCTCTGGACGAGAGCGCGCTTACCGTTTGAAGGGCTTGGCCAGTACCTGGA 428
Db 1728 CCGATTAGACCGCGCGAGAGAGGGCTTCTGCTGACCAAGCTTGGCGCGCGCTTACAGCGCC 1787
QY 429 CGGGTTGGGGGCGACGCGCGCGCGCTTCAAGCGCGCTTATGATGGGCGCATCAAGCT 488
Db 1788 CTTGTTGGCGCGGAGAGAGCGCGCTGAGCGGGGCGCGGGAGCGCGCGCGCGCGCGCG 1847
QY 489 GCGCGCGCGCTGATGCGCGGCTTGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 548
Db 1848 AGATGACGAGAGGGGGTGGCGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1907
QY 549 GCGGAGCTTGGCGCGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 608
Db 1908 CGGGTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1967
QY 609 CTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 662
Db 1968 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2027
QY 663 GCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 722
Db 2028 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2087
QY 723 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 782
Db 2088 GAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2147
QY 783 GCACCAACAAGTATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842
Db 2148 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2207
QY 843 CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 902
Db 2208 GCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2267
QY 903 CACATGCTCAAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 962
Db 2268 GCTGCTCTCATG-----CGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2321
QY 963 GCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1022
Db 2322 CGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2381
QY 1023 CAATAGACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1082
Db 2382 GCGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2441
QY 1083 ACCGAGTGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1142
Db 2442 CCAAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2501
QY 1143 GCAAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1202
Db 2502 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2561
QY 1203 TGGCGCGAGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1262
Db 2562 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2621
QY 1263 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1322
Db 2622 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2681
QY 1323 GGTGTGTTTGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1376
Db 2682 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2741
QY 1377 CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1436

```

```

Db 2742 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2801
QY 1437 GCGGTTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1496
Db 2802 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2861
QY 1497 CCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1541
Db 2862 CTACTGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2906

```

## RESULT 15

```

US-09-815-242-4056
Sequence 4056, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlson, Karl L.
APPLICANT: Zyckind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Travick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4056
LENGTH: 3189
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4056

```

```

Query Match 4.1%; Score 72.4; DB: 10; Length 3189;
Best Local Similarity 43.8%; Pred. No. 1.5e-07;
Matches 371; Conservative 0; Mismatches 471; Indels 6; Gaps 1;

```

```

QY 16 GCGGAGAGAGATTTCATGTCATATCCCATTCAGACCGCTTGCATGCTGCTGCGG 75
Db 319 GTGACGTCGACGAGAACCGCTTACCCGAGACCGAGCTCCGAGAGAGAGAGAGAGAGAG 378
QY 76 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 135
Db 379 CTGGGATATACCGTGCACAAAGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 438
QY 136 GCGGACCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 195
Db 439 CCGGATTAACCGCTACGACATGCTTACCTGTGAGACTAGCGCGGTGCTCAAGAGAGAG 498
QY 196 GGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 255
Db 499 GAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 558
QY 256 CTGCTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 315

```

Db 559 CTGCGCTGTGCTGGACCCGAACAAGTGGCTCGCGCAACCTCACCGCACCGACGTG 618  
QY 316 GGCATTTCGGGATCAGCCCTGGCCGCTGGCGCACGCCAACCCGCCGCCCTTGAG 375  
Db 619 GTCAAGCCGATCCGGAGAAACCGCCAGGTTCGCCCGGCACTCGGGCGCGCCG 678  
QY 376 CTGCTGAGCAAGGGGCTTACGGTTTCAGCGCTTGGCCGACTACCTGAGCGGTTG 435  
Db 679 GCGCCGAGGATACAGCTTCCAGTTGTGATCAACACCAGGTCGCTGTGCACCGAG 738  
QY 436 CGGCGCACGCCCGCGCCCTCAACGCCGCTGTATGTTGGCCATTCAACGCTGCGGCC 495  
Db 739 GAAAGATTCGAACATCATCATCCGCGCGCGGCCAACGGGAGATCACCCGTGCGCG 798  
QY 496 GCGGTCAATGCCGATTTGCAAGCGCGCCGACCCAGCAGAGAAATCGGGCCATCGCGGAC 555  
Db 799 GACATCGCCCGGCTGAGCTGGGCTCAACCAAGTACGCCCTCGTTCGCTGCAACAAC 858  
QY 556 CTGGCCGAGGAAGCATGGCCAGCGGCCCATCGGCATTTCGACCGCGCCTTCTACCG 615  
Db 859 AAGCCGGCGGTGGCAATCCGATCTTCCAGCGTCCCGGCTCGAAGCCCATCGAGATCTCC 918  
QY 616 CCGCGCGCGCGCCGACCCAGCAGATCATGAGTGTGCCGCGCGCTGAGCGCGCAT 675  
Db 919 AACCTGTGGCGGAGAAAGATGGCCGAGTGAAGCACAGCTTCCCGAAGCATGGACTAC 978  
QY 676 GCGGCAATCTACGCCACCCACATGCGCGAGAGGGAGACACATCTGTGGCGCGCTGAG 735  
Db 979 TCCATCGTCTACGACCCGACCATCTTCGTCGCGG-----CTCCATGAGGGGGTGTG 1032  
QY 736 GAAACCTTCGCGATGCGCGGAGCTGGACGTGCGGTGATCTCGCACCAAGATC 795  
Db 1033 CACACCTGTCTGAAGCCCTGGTGTGCTGTGTGTGTATCTGTCTGACAGACC 1092  
QY 796 ATGGCCAGCCCAATTTGCGCGCTGCGCGAGAGCCTGCGCTGATCGAGCGCGCATG 855  
Db 1093 TGGCGCGCGCTGATCATCCGCTGGCGCGGGTGCCTGTCTGATCGGCACCTTCGCG 1152  
QY 856 GCGCGCCA 863  
Db 1153 GTGATGCA 1160

Search completed: May 11, 2003, 12:09:59  
Job time : 447 secs